

Growing with environmental care

The World Bank's annual report on development strategy shows its full commitment to environmental causes. When will that go so far as to link assistance with prudent population policies?

THE World Bank's conversion to the cause of sustainable development appears to be complete, to judge from its *World Development Report 1992*, the fifteenth in an annual series, issued earlier this week. Lewis T. Preston, the bank's president, says the main need now is "to integrate environmental considerations into development policymaking". But the bank's traditional optimism has not evaporated: the document insists that "continued and even accelerated" development can be sustained. Unlike other pronouncements on these issues, the Brundtland report for example, the bank has taken care that its case is sustained by evidence. Indeed, there is such a wealth of data on the nexus between development and environment that those packing their bags for next month's conference in Rio de Janeiro would be well advised to put a copy in their luggage.

Two particular issues deserve attention, of which the most immediate is the positive correlation between prosperity and many important aspects of environmental quality. That is not, of course, a surprise. Only rich countries can afford safe public water supplies, of course, but the data also confirm that urban air is less polluted (by dust and sulphur dioxide) in richer countries. That merely shows that environmental protection and improvement is a consumer purchase (often, of course, by governments on behalf of their taxpayers). The point is neatly illustrated by a graph showing the sulphur dioxide concentration in urban atmospheres as a function of the average per capita income of the countries concerned. The most polluted cities are those in countries with an income of about \$1,200 per head per year. Countries poorer than that produce little pollution anyway, richer countries can afford to clean up the air.

The implications for the future of the World Bank's policies are plain. Development used to be an objective in its own right. With the addition of the goal of environmental protection and improvement, increasing the prosperity of the poorest half of the world becomes an even more compelling need, for otherwise the poor will not be able to afford the measures they need to take. The logic is compelling, simple though it may be. The fly in the ointment is that the goal may be, for many countries, unattainable because the poorest countries include those in which the rate of population growth is greatest. Whatever assistance they are given, the argument goes, they will continue to grow more quickly than the economic resources at their disposal, with the consequence that they will never be able to comply with seemingly environmental requirements.

That is yet another reason why the bank is driven to the emphasis on population growth that marks this week's document. What, other than exhortation, can be done? Commendably, the bank has accepted the only forceful analysis of the problem — that the causes of rapid population growth include the insecurity of poor families whose calculations of their future all too often require them to suppose that half their offspring will probably be dead by puberty. Who can blame peasant families for producing more offspring than are strictly necessary when, in such circumstances, they can only fear that there will be nobody left to execute the most primitive life insurance policy in the world, filial regard for older generations?

So it makes entire sense that the World Bank now plans that a greater share of its resources should go on humdrum projects such as the provision of safe water supplies. Not only are these desirable in themselves, but by reducing the incidence of infant deaths, they may be the quickest way of giving poor populations the sense of security in the future than the rich enjoy. The education of women is the second prong of the World Bank's efforts in this connection — again an intrinsically desirable goal whose side-effects include the abatement of population growth. So far, so good. The snag, so far, is that the bank does not go so far as to say that policies on population growth will be a necessary qualification of requests for development assistance. For how much longer can it persist with that gentility? □

Anti-environment Bush

President George Bush has sided with industry against his own environmental chief on an air pollution issue.

US INDUSTRY won a round in an environmental fight last week when President George Bush decided to allow companies to exceed their pollution permits by as much as 80,000 pounds of toxic chemicals a year without first giving public notice. The president's decision is, perhaps, more important for its symbolic value than for any actual effect on pollution, for the decision is certain to be challenged in court.

The heart of the issue is an argument over the interpretation of the US Clean Air Act of 1990, which the head of the Environmental Protection Agency (EPA) says requires public notice (and with it inevitable cost and delay) from companies wanting to dump or spew into the environment more pollut-



ants that they are allowed. Industries that are particularly affected include large-scale chemical plants, oil and gas refineries, and other petrochemical-based industries. Arguing that public notice of an excess of a mere 80,000 pounds a year constitutes an undue regulatory burden, industry went to the President's Council on Competitiveness, headed by Vice-President Dan Quayle, for support against the EPA.

Even though companies can skip the public notice provision of the law, they will still be required to notify state governments of their intention to exceed pollution permits, and states, in turn, can notify EPA, thereby demanding two administrative steps before excess pollution is finally released into the air. Compared to the millions of pounds of pollutants that industry legally discharges into the environment every year, an additional 80,000 pounds one way or another should not seriously affect the financial well-being of any large company.

Of all the environmental issues before the United States, this should not rank very high. Certainly it should not require a presidential decision for its resolution. By stepping into the argument on the side of pollution, Bush damages his own claim to being an 'environmental president' and sends exactly the wrong message to his international colleagues who will be assembling next month at Rio de Janeiro. Perhaps, after all, he would be well-advised to stay home. □

Muddled carbon tax

The European Commission's proposals for a carbon tax are only a first approximation to what will be needed.

DR Carlo Ripa di Meana is one of the European Commission's most energetic members who, for several months, has laboured on a proposal for a Europe-wide carbon tax. The objective is as laudable as it could be, but is also rational. If atmospheric carbon dioxide reinforces the greenhouse effect and thus is part of the cause of potentially uncomfortable or even disastrous global warming, it must surely make sense to restrict its emission by a tax rather than by other means, say physical controls of who burns how much of what. If there have to be restrictions, it is best that they should take the form (as with a tax) of extra economic costs. Then, it may be presumed, people will make rational decisions about the fuels they use, economizing especially in those that cost them most, and thus collectively minimizing their emissions of carbon dioxide.

It is therefore to be hoped that di Meana will not be too downcast that his proposal, made public after a meeting of the Commission last week, has been denounced by Friends of the Earth as proof that Europe's claims to be environmentally enlightened are "fraudulent posturing". The organization says that Europe "has bottled out and deferred to George Bush". The occasion for this intemperate protest is that the introduction of a carbon tax in Europe has been made conditional on agreement by the governments of the United States and Japan that they will introduce comparable schemes. Not to have included that condition would have been tanta-

mount to European agreement to shoulder on behalf of its trading competitors first responsibility for action against global warming. A commission that took on that onerous responsibility would have been laughed at in the United States and Japan, and probably quickly voted out of office. On that score, di Meana has nothing to repent of.

The real complaint against di Meana's carbon tax is different: in present circumstances, it would be ineffectual. To be sure, the details are not yet settled (but may be at a ministerial meeting planned for next week). Meanwhile, the intention is that there should be two parallel taxes — one on carbon-based fuels and another on energy consumption (based on the calorific value of the source of energy). The hope is that the first tax would persuade energy-users towards fuels contributing relatively little carbon dioxide to the atmosphere, and that the second would generally limit the consumption of energy as such. For starters, the Commission has apparently in mind a tax rate of about \$10 for a barrel of crude oil. To guard against the possibility that, by taking money out of circulation in the middle of a recession, the taxes would turn recession into slump, the Commission proposes to urge on member governments that the extra taxes should be offset by reductions elsewhere.

There are two grounds for rational scepticism. First, there is ample reason why the market for energy should be insensitive to its price. In a cold winter, people use what energy they need to keep warm, and only afterwards worry about the cost. More to the point, a person who has spent, say, \$30,000 on an automobile is unlikely to be strongly deterred from using it by an increase of the price of gasoline (petrol) of two per cent or so when the annual cost of fuel is unlikely to exceed, say, \$3,000. To say that is not to say that the principle of a carbon tax is pointless, but that such a fiscal device would function as intended only if it were pitched much higher. That would hurt, but might mean that some European governments could finance all their operations from their taxes on energy and carbon.

The other objection to di Meana's proposal is that it muddles together the objectives of energy conservation and the reduction of carbon dioxide emissions. The latter is the overriding objective, the former a means to that end. (If solar power were plentiful and cheap, nobody would have to bother about energy conservation.) So why tax them separately? And why, in particular, decree (as di Meana intends) that nuclear power should bear the same energy tax as, say, fossil based electricity generation? The carbon tax should encourage switching between fuels, although inadequately; the accompanying energy tax would blunt its effect by making the economic benefits of switching less marked. Is the penalty on nuclear power just an old-fashioned way of making sure that the French do not get too far ahead?

That is one good reason why the whole proposal deserves a longer and more careful inspection. Everybody can sympathize with di Meana's wish to have something tangible to put on the table at next month's conference in Rio de Janeiro. Half-baked though it is, his carbon tax will at least demonstrate that Europe is ready for action of some kind. It should also encourage the others. But there is an urgent need that it

should be widely regarded as what it is — a first approximation to a workable carbon tax of a kind that will probably be needed one day. Di Meana should be remembered with honour for that. □

Realities behind riots

Blacks are underrepresented among graduate students in the United States. Why should this be?

IN the aftermath of the Los Angeles riots, there has understandably been hand-wringing and introspection about inner cities in the United States. Predictably, there has also been a rash of opinion polls and statistical collations attempting to sum up or point to solutions for these difficult problems — often in the form of over-simple pie charts and histograms. Two such exercises, in particular, stand out because of their common thread — the education and training of black Americans. Thus a *New York Times*/CBS News poll has found that 78 per cent of respondents agree that more jobs and job training would “help a lot to reduce racial tension and prevent riots”.

An admirable goal, nobody would deny, but how likely is it ever to be attained? Hard statistics from the US National Research Council, just published, show that black Americans have a hard time in graduate schools. Among black US doctoral candidates, 63 per cent have to support themselves financially, whereas 69 per cent of graduate students from outside the United States are given research or teaching assistantships (a higher percentage, incidentally, than that for all US candidates, 42 per cent). Such arrangements, some say, ensure that black students are put off graduate studies even before they reach for the application forms. The comparison, of course, is not entirely fair. Foreign students, for example, are more likely to be found in scientific disciplines whereas black Americans favour liberal arts, which are relatively poorly provided with research grants from which to pay graduate students.

Even so, the fact remains that little more than two per cent of PhDs awarded in 1990 went to black Americans and in some subjects — conspicuously molecular biology and applied mathematics — no black student was awarded a doctorate that year. Even in a subject such as education, the proportion of blacks being supported was much smaller than the proportion of foreign students. Any attempt to divine the underlying reasons soon runs into the quicksands of subjectivity and racial prejudice, encapsulated as “the irrational fear of black men on the part of those in charge of graduate programs” and “the unnerving image of aggressive black males in the laboratories”.

What is to be done? Have two decades of positive efforts to end discrimination failed to exorcise these prejudices? Should not the clearly increasing interest of social scientists in the image projected by young black men (in particular), and the misreading of that image that is all too common, be channelled into studies of what goes on in the admissions departments on their own doorsteps? While waiting for

public high-school education to be put to rights, making sure that qualified black students make the most of what educational opportunities there are would be a prudent stop-gap. □

DNA feud

Apparently the US EPA cannot agree with other agencies on the regulatory status of recombinant DNA.

A CONSENSUS has been reached in the scientific community generally that safety concerns about agents produced with recombinant DNA technology should be directed to the characteristics of the recombinant DNA agent in question rather than to the process by which it is made. Even the White House says it agrees (see *Nature* **349**, 726; 1991).

Nevertheless, the US Environmental Protection Agency (EPA) still bases regulatory decisions on process rather than product, according to a letter from a University of California (UC) scientist and EPA adviser who has challenged EPA's official statements that it is scientifically valid to regulate the process of recombinant DNA technology. Robert Burris of UC, Riverside, says he thought it was a “political decision of convenience, rather than a scientific one” to go on pretending that DNA technology is potentially hazardous in and of itself. Burris is right and EPA should see the scientific light. □

The great fungus

A killer fungus from Washington State challenges benign *Armillaria bulbosa* for the title of world's largest mushroom.

A MONTH ago, two botanists from the University of Toronto and a forester from the Michigan Technological University reported the identification of a fungus in the state of Michigan said to be among the ‘largest and oldest living organisms’ in the world (Smith, M. L. *et al. Nature* **356**, 428; 1992). Weighing more than 10,000 kg, and occupying a minimum of 15 hectares or 38 acres, the 1,500-year-old, genetically stable *Armillaria bulbosa* certainly sounded like a world-class giant. But now a competitor for the title has reared its nasty head.

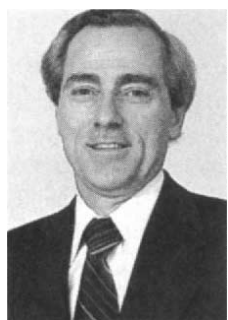
A pathologist at the Department of Natural Resources of the State of Washington, in the northwest of the United States, says that Washington has the world's largest organism — an *Armillaria ostoyae* that covers 1,500 acres in the foothills of Mount Adams. But *A. ostoyae*, whose age is merely estimated at 400 to 1,000 years, is a youngster by comparison with *A. bulbosa*. It also sounds like a less friendly creature: *A. ostoyae* is a killer fungus that can wipe out whole populations of trees. Why Washington should be so cursed, while Michigan's fungus is benign, remains unexplained.

But that is not an end to the affair. The word is that Oregon State may also enter the competition. Who knows where it may go from there? May the best fungus win. □

NIH scientists chafe under ethics rules on industry ties

Washington. You are a scientist from the US National Institutes of Health (NIH) on a visit to a company. Someone calls a lunch break and offers you a tuna-fish sandwich. What do you say? "Thanks, but I'm not hungry."

That, at least, is what NIH scientists are now being taught in a training video-tape. The tape, along with almost weekly memoranda on other issues involving contacts with outsiders, has become part of a body of ethics rules and regulations that threaten to turn the agency, once renowned for its research freedom and hassle-free science, into one of the most restrictive in the government.



French Anderson paid the price for blazing trail to industry.

The culprit is the 1986 Technology Transfer Act, which encourages US agencies to collaborate with industry to commercialize government inventions. One of the mechanisms for this is a Cooperative Research and Development Agreement (CRADA), in which government and industry scientists work together with the ultimate aim of generating a marketable product. NIH now have about 160 CRADAs, more than any other agency, and it is that success that worries NIH administrators.

The problem is that collaboration means frequent visits between NIH scientists and industry, each one of which, from an ethics officer's point of view, offers an opportunity for undue influence and profiteering. Scientists who consult and give speeches outside their usual NIH duties, as more than 440 NIH researchers did in 1990, are also seen to be skating on thin ethical ice.

With congressional investigators ever vigilant, and science itself under unprecedented scrutiny, NIH officials have become increasingly sensitive to even the appearance of conflict. The result is some of the most intrusive guidelines in the government. Starting from government-wide regulations issued by the US Office of Government Ethics in 1989, NIH officials have tried to respond to each new ethical dilemma raised by the agency's expanding industry connections.

The evolving rules and guidelines range from the cumbersome to the absurd. One memorandum cites the example of an NIH scientist who had been given approval to give a speech as an 'outside activity' — something that is not strictly a part of his

official capacities. But the researcher made the mistake of discussing his published work at length. According to the memorandum, he should not have talked about his own work for more than three minutes of every hour. Any more, NIH decreed, smacks of using his "public office for private gain".

Another memorandum tackles an apparently trivial aspect of privately subsidized travel: the free shaving kit that some airlines offer. If it is worth more than \$5, the kit might qualify as an inappropriate gift. Better not to accept it, NIH ethics officers warned.

The entire issue of travel is sensitive since Congress slashed the travel budget of NIH's parent agency in retribution for sending what some legislators thought were too many researchers to last year's International AIDS Meeting in Florence, Italy. Now, with little money to travel on their own, NIH scientists must often wait months to gain approval to accept airfare for outside activities, and requests for paid consulting must go all the way to the NIH director.

"For all practical purposes, it's a moot point", says Jeffrey Schlom, chief of the Laboratory of Tumor Immunology and Biology and Experimental Oncology at NIH's National Cancer Institute. "If a company called here tomorrow to ask me to visit them, I couldn't go. There's zero money in our travel budget and it takes months to get approval."

As the first NIH scientist to become involved in a CRADA, W. French Anderson can offer more than the usual number of horror stories. A pioneer in genetic therapy, Anderson has several CRADAs with Genetic Therapy Inc. (GTI). To protect himself from what he assumes are the inevitable questions, Anderson has chosen to rearrange his life. His salary payments are sent directly to his lawyer, who controls his bank account and his investments. He says he has spent about \$50,000 of his own money on legal advice, and at one point had three personal lawyers. He owns no stock in any company.

When GTI went public, Anderson worried whether his discussions with friends about his work might violate the insider-trading prohibitions of the US Securities and Exchange Commission. At the same time, Anderson became concerned that a federal rule that prohibits a company from advertising its connections with government researchers might clash with another rule that a company must identify its affiliation with outside scientists. And earlier this month, when Anderson announced that he might be leaving NIH and taking an academic position to follow his wife to the West

Coast, he found that he was prohibited from even discussing the future of his collaboration with GTI. Only a waiver last week from the NIH director, Bernadine Healy, allowed him to begin planning how to keep the collaboration alive.

Even after the lawyers had worked out the ground rules of the collaboration with GTI to avoid any possible perception of conflict of interest, Anderson thought of something that they had missed. Every Tuesday, when he visits the nearby company, tradition has it that he gets a chocolate doughnut. Would this be the improper 'gift' that finally brought in the investigators? Better safe than sorry, he decided, giving the company \$50 to establish a doughnut fund for him. "To live your life on the advice of a battery of attorneys is no fun," he says.

Patricia Werner, NIH ethics coordinator, explains that the rules are designed with the researchers in mind: "The whole idea is not to keep scientists from doing what they want to do, but to protect them" should they be investigated, she says. But many NIH researchers would rather accept the risks. "These rules seem totally illogical and repressive", says Schlom. "I don't think the ethics people realize what the procedures have done to day-to-day life at NIH."

"It's just a very cumbersome process," agrees Philip Chen, NIH associate director for intramural affairs. He believes, along with many of his colleagues, that NIH has overreacted to the federal guidelines. But he cannot see how to ease restrictions without inviting in crusading investigators. "We're in the early stages of developing a technology transfer process", he explains. "I admit that there's an awful lot of red tape, but I don't see any good way to reduce it at this moment."

The situation may improve next month when the Office of Government Ethics is expected to release its new standards for ethical conduct. A proposed version, issued for public comment last July, sought to clarify many of the issues with which NIH's own ethics officers have been wrestling.

Some of the changes are obvious improvements (the shaving kit, for example, would be acceptable). Others, including a controversial section that would prohibit government researchers from working for their professional societies on government time, have been delayed for further consideration. Although the final language is not yet public, the rules are expected to make unnecessary some of the most extreme agency regulations. But, given the current climate of fear at NIH, few researchers expect a return to a time when they were free to concentrate on science.

Christopher Anderson

French propose agency to spur European defence research

Paris. The French minister of defence, Pierre Joxe, has called for the creation of a European Armaments Agency (EAA) to coordinate big-science projects in military research. Speaking last week at the close of a two-day conference entitled "Science and Defence: New Europe, New Research", Joxe also announced research collaborations with the Commonwealth of Independent States, the creation of a French version of the Massachusetts Institute of Technology (MIT) and new ways to integrate French defence and civilian research programmes.

European countries already cooperate in military research. In 1989, 13 of them launched the EUCLID (European Cooperation for the Long term In Defence) programme to foster multinational cooperation in fundamental research. The programme supports work in areas related to simulation, microelectronics, acoustics and radar, robotics and artificial intelligence. But decisions about who will participate, and how much each government will contribute, have slowed progress, and it will be several more weeks before the first contracts are announced.

No details are available on how the EAA would operate beyond that it would be modelled after the European Space Agency. Joxe imagines that it would begin by assum-

ing management of EUCLID and other multilateral research agreements.

One example of the type of research that could be carried out under EAA is a proposed telecommunications satellite programme. The French Syracuse II telecommunications satellite is a joint venture between the Ministry of Defence and France Telecom. But the rising cost of such satellites has already led to discussions of cooperative efforts by a newly formed group, EUMILSATCOM, that represents Britain, France, German, Belgium and the Netherlands.

French officials insist that the EAA poses no threat to the national security of individual countries. "As the interests of each country become more common, they will share secrets", says Jacques Lanxade, the army's chief of staff. "France is ready to share." Joxe says that 90 per cent of French military research, excluding nuclear weapons, is already in the public domain.

The new armaments agency would not prevent research agreements between individual nations. France has made an initial investment of FF50 million (US\$10 million) towards research cooperation with the countries of the former Soviet Union, and defence officials have begun to evaluate the capabilities of laboratories in those coun-

tries with a view to joint research, licensing, scientific exchanges and subcontracting.

Internally, France hopes to take advantage of whatever 'peace dividend' exists to increase research spending in such areas as biomedicine, space, information technology and communications, acoustics and robotics. This year's defence research budget will nearly double funding for space research, to FF450 million, and will significantly increase support for research in public laboratories and universities. It comes at a time when almost every other part of the military budget is being slashed.

Joxe and Hubert Curien, minister for research and space, say that they want to "tighten the links between civil and defence research". To that end, the two are considering a centre to develop and test advanced robots with dual uses, and an announcement is expected soon that will add military objectives to the role of the space research agency Centre National d'Etudes Spatiales.

The hub of France's proposed version of MIT will be the Ecole Polytechnique at Palaiseau near Paris. One of the oldest and most renowned of the Grand Ecoles, the *polytechnique* trains engineers and scientists for the nation's high-technology companies. A science and technology *pôle* (centre of excellence) (see *Nature* 356, 373; 1992) will be formed at the site, bringing together the Ecole Nationale Supérieure des Techniques Avancées, the government's weapons laboratories, the army documentation centre and a proposed technology transfer centre to be called *X-pôle*.

David Bakewell

Swiss vote for biotechnology, with controls

Basel. Switzerland's first law regulating reproductive medicine and biotechnology has been written into the statute following its approval by 74 per cent of the population in last weekend's plebiscite. The law provides a framework for more detailed regulations, and pressure groups are already campaigning for tighter controls on biotechnology that could have severe consequences for Switzerland's drug companies.

In Switzerland, any proposal to introduce new legislation must be taken to a public vote if it gains at least 100,000 signatures. The government has the right to put forward a counter-proposal. Either measure becomes law if it receives a majority vote.

Prompted by the absence of any specific legal control of genetic engineering, pressure groups led by the political magazine *Der Beobachter* collected 127,000 signatures in favour of a draft code in 1987. But they withdrew it after endorsing the government's counter-proposal.

It is this counter-proposal that passed on 18 May. Swiss scientists interpret the vote as a public endorsement of genetic engi-

neering, subject to control. It defines clearly new regulations for reproductive medicine. *In vitro* fertilization and embryo transfer will be restricted. And the use of surrogate mothers, manipulation of embryos (for example, to determine sex) and the introduction of non-human genetic material into the human genome are expressly forbidden.

The new law is less clear about biotechnology, suggesting that the government should act "to ensure the dignity of organisms and the security of man, animal and environment". Bernhard Bühler, spokesman for the Swiss Society of Chemical Industries, says that this language conforms to the ethical guidelines laid down by the society two years ago.

But a different set of pressure groups, under the umbrella of the Swiss Working Group on Genetic Engineering (Schweizerischer Arbeitsgruppe Gentechnologie, SAG), are already campaigning for much stricter federal regulations. Twenty-three organizations, ranging from the World Wide Fund for Nature to the Association of Small Farmers, collected signatures in front of last

Sunday's polling booths. Their proposal would ban both the patenting and the release of genetically manipulated organisms, as well as the creation of mutant animals. And it promises very strict regulations, as yet undefined, on industry.

Basel's powerful anti-biotechnology lobby already influences the decisions of its local pharmaceutical giants. Last December, Ciba-Geigy withdrew its plan to build a large biotechnology centre in Basel; it is now being built 10 km away, in France, where the political climate is better and there are fewer bureaucratic obstacles. Drug companies are concerned that the situation would become even worse if the SAG manages to place its proposals on the ballot.

Paul Walter, president of the Swiss Academy of Science, says that it would be hard for industry to cope with the provisions of the SAG initiative. But he sees the large margin by which the compromise was accepted (support in Basel was 82 per cent) as a sign that Swiss voters are reluctant to do anything that would threaten the country's economic health.

Oliver Klaffke

Germany to shift funding from physics to biology

Munich. After a decade of generous support of large-scale physics programmes, the German Ministry of Research (Bundesministerium für Forschung und Technologie, BMFT) has decided to direct more of its funds into centres of excellence in the biological sciences. The change will relegate to a lower priority such new large items of equipment as accelerators.

The BMFT has adopted in principle the newly published recommendations of its advisory committee, chaired by physicist Siegfried Grossmann. In addition to a restructuring of Germany's research base, the committee has also suggested that industry should put more money into basic research, in particular such areas as chemistry.

The Grossmann committee was created in July 1990 by Germany's research minister, Heinz Riesenhuber, to analyse the balance of BMFT research spending and to recommend changes in line with Europe's new political map. Germany's spending on research, at 2.9 per cent of the country's gross national product, is already one of the highest in the world. About a fifth of that is spent on basic research. But the unexpectedly high cost of reunification, as well as a creeping increase in the proportion of BMFT funding directed to basic research (from 26 per cent in 1982 to 40 per cent in 1990) has forced a rethinking of the country's research priorities.

The Grossmann report stresses the need for a greater commitment to the life sciences. At present, only 14.5 per cent of the BMFT budget is invested in biomedical research, including biotechnology. Committee member Horst Kern, professor of cell biology at the University of Marburg, says that the United States carries out about 80 per cent of the world's important molecular

biology research. A greater, long-term commitment to biomedical research that exploits new biological technologies would help strengthen the German economy, he says.

The report proposes new centres to specialize in genetics and neurobiology. They would be analogous to the National Cancer Centre in Heidelberg, the ministry's major recipient of biomedical research funding and renowned for its work establishing a link between papillomavirus and cancer of the cervix. It also recommends greater support for existing groups at universities and research institutes working in such areas as AIDS and infectious diseases. Harald zur Hausen, director of the Heidelberg centre, says that the absolute level of biology funding has long been too low. He hopes that tight budgets will not prevent the ministry

from carrying out the recommendations of the Grossmann report.

Despite the new emphasis on the life sciences, Riesenhuber says that large current projects in the physical sciences are unlikely to be curtailed. But requests for expensive equipment, he says, are more likely to be turned down. Speaking for the BMFT, Christian Uhlhorn says that the ministry believes that it is now time to assimilate data from its investment of billions of marks on large-scale equipment research in the physical sciences rather than replace machinery with newer models. Such a policy, for example, could limit Germany's contribution to the proposed Large Hadron Collider at CERN, the European Laboratory for Particle Physics in Geneva.

Final decisions about the fate of various physics projects await the announcement in July of the 1993 BMFT budget. The new budget is expected to increase the ministry's current allocation of DM9,200 million (US\$5,000 million) by only the rate of inflation — an amount unlikely to be enough to cover the research demands of a unified Germany.

Alison Abbott

Germany seeks new funds to meet costs

Munich. Germany's science research council wants to abandon its promise to accept an annual increase of only 5 per cent in its funding because it needs more money to cope with a sharp rise in grant applications and a new wage settlement of 5.4 per cent with public-sector unions.

An agreement was reached in 1990 between the government and the Deutsche Forschungsgemeinschaft (DFG) that would limit growth in the research council's funding to 5 per cent a year for five years. But the DFG president, Wolfgang Fruhwald, announced in January that the so-called 'five-times-five' agreement was no longer tenable. Grant applications had been increasing at an unprecedented level — around 9 per cent a year since 1989 — and, for the first time, applications with high priority ratings were not being funded.

About 75 per cent of the DFG budget is spent on salaries. As most DFG employees work in the public sector, this year's budget increase of 5 per cent will be wiped out by the recent wage agreement.

Fruhwald is pressing for an 8.5 per cent rise in 1993, but DFG says that even that level would be insufficient to cover its basic funding needs. "Twelve per cent would be more realistic", says Eva-Maria Streier, speaking on behalf of the council. Its decision to seek a smaller amount, she says, was intended as a gesture of goodwill.

There are several reasons for the rise in applications, which last year exceeded 10,000. Part of the cause is reunification, although a special fund of DM90 million

(US\$50 million) was created in 1991 to finance more than 1,000 grants for the states in eastern Germany. In addition, the cost of reunification has reduced what other funding sources, particularly the universities and the Ministry of Research, can afford. And 'baby boom' scientists, now qualified researchers wanting to set up their own projects, have swollen the pool.

Horst Kern, professor of cell biology at the University of Marburg and a member of the council's scientific advisory committee, says that the DFG is in a critical position. "For the first time we've reached the situation like that faced by the United States and some other European countries, where highly rated work is not funded."

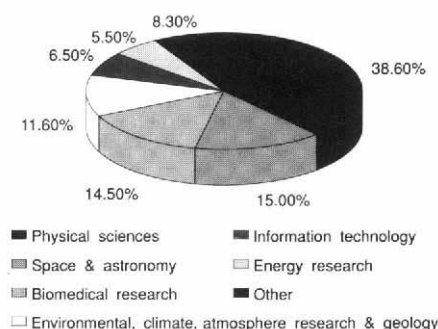
For example, last autumn's meeting of the DFG approved 29 of 44 applications for the priority programme of grants, which fund large interdisciplinary projects and can be worth up to DM6 million. However, the council could afford to fund only 17, and non-scientific criteria such as 'urgency' were brought into the final selection process. Official figures indicate that less than 45 per cent of grant applications in western Germany are now successful.

The council's major source of funding, the German Ministry of Science and Research (Bundesministerium für Forschung und Technologie), will respond to the request for increased funding after its own budget is set by the Minister of Finance in July. Negotiations are also continuing with individual states, which provide a quarter of the council's funding.

Alison Abbott

Allocation of the 1990 BMFT budget for basic research

(total is DM3,170 million).



India goes to sea for research funds

New Delhi. Indian scientists want to turn the blood of the horseshoe crab *Limulus polyphemus* into a lucrative business whose profits can support all areas of science in the country. But US companies already working in the field say that the Indians have greatly overestimated the world market for the product, and they predict that such a windfall is highly unlikely.

The blood of the horseshoe crab is the source of amoebocyte lysate (AL), a chemical that can be used to test for bacterial toxins. It is used by pharmaceutical companies to demonstrate the safety of their injectable drugs and of medical devices that come into direct contact with the bloodstream, such as heart valves or balloon catheters. The test is more sensitive than the previous method, which involved rabbits, and takes only three hours instead of three or four days.

The marine creatures can be bled and returned unharmed to their habitat. A single crab can supply 20 ml of blood that can yield 5 ml of AL. That amount of reagent sells for US\$25 or more, depending on how it is formulated and packaged.

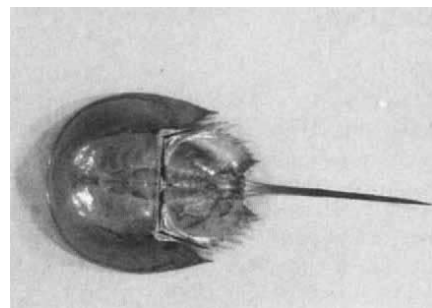
The National Institute of Oceanography (NIO) has set up a facility in Orissa, along

India's eastern coast, to extract AL. It is asking the government for money to scale up the operation to a commercial size. The supply appears bountiful: unlike its American cousin which breeds during the late spring and summer, the Indian crab breeds all year round. The horseshoe crab is also found along the northern shores of Japan and on the southern tip of China.

NIO scientists believe that there is no limit to the amount of crab blood that India can supply to the world's pharmaceutical industry. They talk of a market of US\$5,000 million, and profits that could finance the country's entire research budget.

That view is "wildly optimistic", says Michael Dawson, assistant director of Associates of Cape Cod, Massachusetts, the world's largest producer of the reagent. Although Dawson would not disclose the annual sales of the privately held company, he says that he doubts very much if the world market is 1 per cent of the amount being discussed in India.

"I think that the market is much more mature than they would like to believe", he says. Although demand for the product has risen steadily since the company was founded in 1974 by Stanley Watson of the



Horseshoe crab's blood seen as source of revenue for research by Indian scientists, but US producers question size of market.

Woods Hole Oceanographic Institution, Dawson says that the limiting factor is demand, not production capacity or the supply of crabs. "If people wanted to buy 10 times as much reagent, we could sell it to them", he says.

But Indian scientists are not deterred by such sombre forecasts. They believe that the market for their product, especially in the developing world, is growing rapidly. And they hope to persuade the government that theirs is a gamble well worth taking.

K.S. Jayaraman

Marine blasts to be regulated following UK protest

London. Procedures for setting off marine explosions in the name of research are likely to change following recent protests by the environmental group Greenpeace over work being carried out by the British Geological Survey (BGS). Although Greenpeace's objections did not stop a detonation off the coast of Wales — bad weather had already caused it to be postponed — the ensuing publicity drew attention to the fact that there are no formal procedures for regulating such work.

The BGS's marine blast was planned for Cardigan Bay over the weekend of 9–10 May. It marked the final stage in a study of the deep basement structure 6–8 km below the surface of Wales. A series of land detonations had already been set off without complaint.

Although the Geological Survey insists that its informal procedures for assessing the environmental impact of marine explosions are more than adequate, it says that subsequent studies will require a written protocol, including an assessment of possible effects on marine wildlife. Greenpeace was particularly concerned that the blast could harm bottlenose dolphins and harbour porpoises.

Although BGS will probably implement new procedures, the Department of the En-

vironment has no plans for legislation on the matter. One reason is that a legislative solution would probably have to cover marine explosions carried out by the military as well as those by researchers. Much less is known about military detonations, but both the geologists and Greenpeace agree that these are much larger than those used in geological work.

In response to Greenpeace's protest, BGS prepared a detailed written protocol for the blast. Although BGS believes that the protocol differs little from the conclusions reached by its informal methods, additions include a diver to ensure that the charge is properly in place and a limit of 25 kg of explosives. The original plan called for a charge of as much as 50 kg if sea conditions were rough to make sure that the vibrations would not be dissipated.

The marine blast was finally postponed earlier this week when the group responsible for monitoring sea mammals in the area — the Dyfed Wildlife Trust — withdrew from the project. BGS was unwilling to go ahead with the detonation without suitable monitoring, and is blaming pressure from Greenpeace for the group's withdrawal. The blast will go ahead at a later date, but reassembling the monitoring equipment will cost several thousand pounds.

Greenpeace has said that if it is unhappy with the situation in the future it will intervene again.

Although geologists say that the written protocols are no great burden, they worry that whatever measures are taken will not satisfy groups such as Greenpeace. At the same time, marine biologists welcome the increased scrutiny being given to such detonations.

"It is quite unacceptable that people go about letting off charges without following some sort of guidelines," said Christine Lockyer of the Natural Environment Research Council's Sea Mammal Research Unit. Although researchers from the unit were consulted as part of BGS's informal assessment, they were not entirely happy with the situation. "We were given no details of how the charge was to be set off or how much explosive would be used", said Lockyer.

In the past, protests by Greenpeace that detonations used in oil exploration were killing fish led oil companies working in the North Sea to abandon explosives in favour of air-gun arrays. Geologists have also adopted air guns, which are supposedly less damaging, but the guns are not sufficiently powerful for deep structure studies.

Ian Mundell

Cabinet shifts unlikely to help South African universities

Cape Town. University officials see the appointment next month of the fourth minister of national education since September 1989 as an unfortunate sign that the government of Mr F. W. de Klerk remains unwilling to make the type of reforms needed to ensure that the country has enough scientifically trained workers.

Piet Marais, who as the present Minister of Education and Culture oversees the country's white education system, will be responsible in his new post for funding university research and the research councils as well as for overall educational policy. But observers say that he is no more likely to deal aggressively with the problems facing the country's education system than was his predecessor, Louis Pienaar. Experts estimate that the country must train 125,000 new teachers and improve the skills of 84,000 others by the turn of the century while integrating the country's racially structured education departments into one system.

"If the universities are to do anything other than muddle along, there has to be bold intervention", says Hugh Amooore, registrar of the University of Cape Town.

The universities themselves have been unwilling to come to grips with the impact of the twofold expansion in student numbers of the past decade. Graduate numbers in science and engineering have dropped from 17 per cent of the total to only 13 per cent, and those in medicine from 11 per cent to 7 per cent, reflecting the fact that most African students do not study mathematics and science at school for lack of teachers.

Although there is consensus among government, the African National Council and the private sector that South Africa's manpower needs are not being met by its current supply of graduates, none of the groups has endorsed a proposed solution to the problem that would subsidize fixed numbers of places in each faculty at each university.

Michael Cherry

Botswana needs science help to save plant and animal species

London. Botswana faces a situation that is common to many developing countries — not enough scientific talent. But a group of visiting environmental scientists, believing that the country's need to protect and preserve its plant and animal resources must take precedence over the government's wish to develop its own expertise, are urging Botswana to hire outsiders as a temporary solution to the problem.

In a preliminary report of their findings, members of the mission to the southern African nation said last week that the dearth of professionally trained wildlife scientists and managers is responsible for a lack of environmental and wildlife data on the country. This could hinder implementation of the government's ambitious National Conservation Strategy and also weaken its ability to respond to those who have criticized its record on conservation. The fact-finding mission was commissioned to investigate just such criticisms.

"The immediate need is such that it is not feasible to wait until the current Botswana junior echelons have built up the necessary experience," said Andrea Pfister, an ecologist with the Conservation Foundation in London, who was part of the mission. "The country should request and allow the donor community to provide and/or finance such executive cadres, not on short-term con-

tracts but with a measure of continuity that will ensure that their successors inherit a well-functioning administration, rather than desultory of overdue projects."

This approach would also help to nurture indigenous talent by producing a range of young researchers with formal degrees, preferably based on research work done in Botswana. It would also provide them with enough work experience to give them confidence and authority in making decisions.

The fact-finding mission was prompted by internal and international criticisms of the country's conservation and environmental policies. They grow out of allegations by the film-maker Rick Lomba, a resident of neighbouring South Africa, that the government has colluded in the destruction of the Botswana's unique environment, which includes an extensive wetland around the Okavango delta, and has failed to protect the country's wildlife.

The mission concluded that Botswana's national conservation strategy is "by any standards a remarkable and impressive document. Few countries can boast such a strategy and the consultation process which produced it." But the real test of the government's commitment, it said, rests on the implementation of that strategy — and the talent to carry it out.

Ian Mundell

IN BRIEF

Washington. Prompted by Congress to think small, the US National Aeronautics and Space Administration (NASA) has drafted plans for a series of cheap, quick planetary probes for later in the decade. Each mission would cost less than \$150 million and be launched within three years of its initiation. In a report submitted to Congress earlier this month, NASA said that the new approach would make space accessible to more investigators and institutions and fill the gaps between the larger missions already planned.

Modelled on the existing Explorer and Earth Probes projects, the new Discovery programme would include a proposal to place a total of about 16 small automated surface stations on Mars to study the planet's geology and weather. The European Space Agency is studying a similar proposal, NASA said. **CA**

Washington. Seymour Cray, the inventor of the supercomputer, has finally agreed to consider what most other supercomputer manufacturers already believe to be the future of the field. Cray, the chairman of the Cray Computer Corporation, told an annual stockholders' meeting last week that he was "having discussions" with makers of massively parallel supercomputers about a joint marketing arrangement. Such an agreement would mark Cray's first radical departure from the design he pioneered more than 20 years ago. Massively parallel machines achieve their speed by combining hundreds or thousands of simple and cheap processors, in contrast to the small number of very fast and expensive processors in Cray's designs.

Cray's change of heart is partly due to his company's shaky financial position: its unfinished prototypes do not work properly and its only customer dropped out last year. Even so, a move into massively parallel computers does not guarantee financial success. Cray Research Inc., the company that Cray founded in 1972 and left in 1989 to form Cray Computer, recently announced its own shift into the field (see *Nature* **356**, 95; 1992). But last week Cray Research announced that increased competition and ebbing demand are expected to lower its earnings by 10–20 per cent compared with last year. **CA**

Washington. After months of speculation that Nobel Laureate David Baltimore would be leaving the Rockefeller University, Baltimore confirmed early this week that he will depart, but not until 1994. Once the Massachusetts Institute of Technology finishes its new biology building, Baltimore will return to continue his AIDS research at the university where he was a professor and founding director of the Whitehead Institute until 1990. Until stepping down to take a faculty position last December in the wake of a divisive scientific misconduct investigation, Baltimore was the president of Rockefeller. **CA**

Choosing the limits to life

P. H. M. Lohman, K. Sankaranarayanan and J. Ashby

How can the world's population be made more healthy? Study of the age distribution of diseases and causes of death in individual countries will enable nationally appropriate research and public-health priorities to be set.

THE ultimate goal of medical research is to maximize the health of people through the treatment and prevention of disease and to minimize or eliminate exposure to agents that may adversely affect health and well-being. Illustrative examples are the discoveries and use of vaccines and antibiotics to combat infectious diseases; the increasing possibilities for treatment and prevention of cancer and coronary heart disease; and the identification of actual and potential carcinogens in the environment.

Although the goal of a world without disease is not practicable, over the past century the average life expectancy of human populations has increased dramatically. This progress has been paralleled by changes in disease spectra and in the relative importance of different causes of death, each of which differ between countries and populations. Nevertheless, there is a limit for curing fatal diseases which is set by the maximum average lifespan of our species. The question of whether this limit is a biological or a practical one is debatable, but the limit itself has profound implications for defining preventive health measures as well as for setting research priorities. We suggest here that as populations asymptotically approach the maximum average lifespan, health initiatives will result only in a shift in the relative importance of individual causes of death. Thus, attempts to extend the average lifespan under such conditions will have only a marginal effect (or none at all) on the average well-being of the population.

Maximum lifespan

The fact that disease spectra differ between populations is well known. The statistics of WHO for the early 1980s are illustrative (for review see ref. 1). In the developing (non-industrialized) countries, the main causes of death, in per cent, are: respiratory diseases (21), infectious and parasitic diseases (18), cardiovascular diseases (CVDs, 16), perinatal disease, (7), cancer (6), accidents (5) and all other diseases (27). In the developed (industrialized) countries, the profile is different: 48% of all deaths are caused by CVDs, 19 by cancer, 7.5 by respiratory disease, 7 by accidents, and 18.5 by other diseases. But these estimates conceal the fact that the absolute number of deaths in each category is greater in the developing than in de-

veloped countries, as 78% of all deaths in the world occur in the former. Although there has been substantial improvement in average lifespan, there has been no significant change in the average maximum attainable lifespan of our species, which is estimated to be about 95 years²⁻⁴.

In developed countries, the average life expectancy at birth has stabilized since the 1980s. Data for the US population^{2,3} are shown in the table and lead to the following conclusions: (1) the

US AGE TRENDS			
Year	AVA No of years	AVR	LAW
1900	58.3	36.7	27.0
1920	63.7	31.3	21.6
1940	69.1	25.9	16.2
1960	74.5	20.5	10.8
1980	79.9	15.1	5.4
2000		9.7*	0.0

Data refer to 50 per cent of the US population, assuming an estimated maximum human lifespan of 95 years. Modified from refs 2,3.

*Extrapolated by curve fitting.

average population age (AVA, column 2) has significantly increased from about 58 years in 1900 to about 80 years in 1980. (2) On the assumption of an average maximum attainable lifespan of 95 years, the difference between this maximum and AVA in any one time period (average remaining lifespan or AVR, column 3) has progressively decreased. The rate of this decrease (the slope as defined by the fraction of people greater than 50 years old) is approximately linear. (3) Extrapolation of this trend (extension of the slope to 95 years) permits an AVR estimate of about 9.7 years for the year 2000. Unless, in future generations, a 'drop dead' situation occurs (that is, all people live to the maximum attainable lifespan and then die), levelling-off of the AVR to between 5 and 10 years seems both likely and biologically plausible. Such a relatively stable AVR can be considered as a reflection of a state of dynamic equilibrium between different causes of death. (4) The number of years potentially available for improvements in lifespan (life-adjustable window or LAW, column 4), which is the difference between the AVR value and 9.7 (see above), has decreased from 27 years in 1900 to 5.4 years in the 1980s.

The LAW estimates discussed above use 50% of the population as a baseline.

For the total population, similar estimates can be obtained by expressing LAW as a percentage of the respective AVR values ($LAW = (LAW/AVR) \times 100\%$), assuming a linear decrease of AVR in people more than 50 years old. The results of this calculation are shown in Fig. 1. It is clear that with diminishing AVR there is a concomitant decrease in LAW, and as average lifespan of the population approaches the maximum, attempts to affect LAW become progressively unrewarding.

An attempt to compare different populations using this approach is shown in Fig. 2. Here, the curves represent different populations: A is the situation in a hypothetical developing country in 1991; B, that in a developed country in 1991; C that in a developed country in the year 2000; and D, the theoretical concomitant 'drop-dead' situation (the population for which the average age at death is the same as the maximum average lifespan of 95 years).

If the concept that the average lifespan of a population is a consequence of the existence of a dynamic equilibrium between different causes of death is valid (the only invariable being the inevitability of death), curves B and C suggest that our ability to influence the occurrence of a particular cause of death in the populations of the industrialized world is small and will diminish rapidly with residual attempts to increase the average lifespan. In the situation repre-

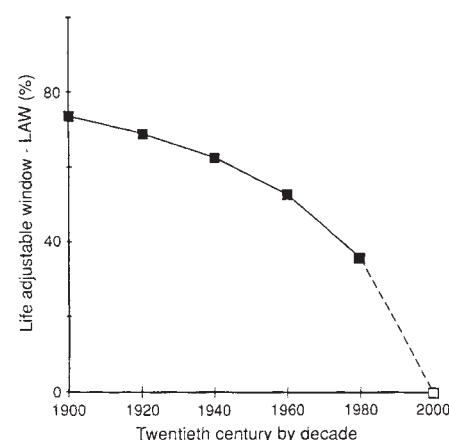


Fig.1 Impact of maximum lifespan on LAW(%), arbitrarily set at zero for the year 2000. Calculations based on linear extrapolation of the curves for each time period towards the average maximum lifespan of 95 years.

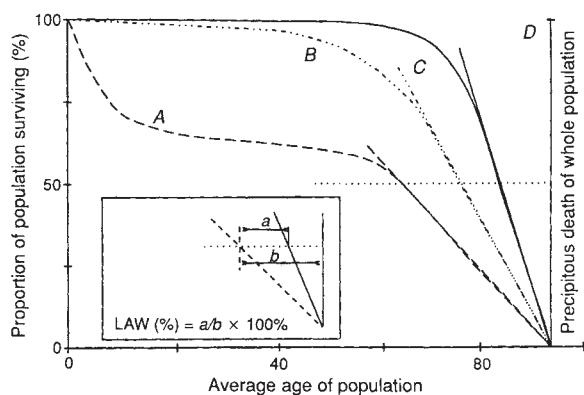


Fig. 2 Hypothetical survival curves for different populations. A, Non-industrialized country 1991; B, industrialized country 1991; C, industrialized country 2000; D, 'drop dead' situation. All curves linearly extrapolated to the estimated maximum average lifespan of 95 years for the year 2000. LAW(%) for a population is estimated as a ratio of distances to curves C and D and in the example shown in the inset for population A. *a*, Distance between the linear parts of A and C; *b*, Distance between the linear parts of A and D. LAW(%) for population A = *a/b*.

sented by curves B and C, for example, where cancers and CVDs are the principal causes of death, significant progress in treatment of CVDs may merely allow geriatric cancers and other degenerative diseases of old age such as Alzheimer's disease to emerge as principal causes of death. Further, attempts to reduce cancer deaths by chemotherapy may result in the induction of secondary cancers: average life expectancies may show a marginal increase but there will be an accompanying increase in cancer mortality figures. In these situations, therefore, it will be possible to 'select' the ultimate cause of death, or at least influence it; quality of remaining life, rather than its length, being the primary variable. The implication that regulators and clinicians will be able to formulate measures actively to 'select' major causes of death in a population is certain to engender much discussion of ethical values.

By contrast, the situation depicted in curve A (non-industrialized country), where a significant LAW is evident, offers ample constructive opportunities for preventive medicine and public-health initiatives. But progress is likely to be slow in view of the much greater population sizes and severe limitations of economic resources. Given the relationship between average life expectancy and the dynamic equilibrium between different causes of death, elimination (or large reduction) of one cause of death will establish a new equilibrium between the remaining causes; elimination of one cause will be offset by deaths due to others, but at a later stage.

Environmental carcinogens

We shall use toxicology as an example of how we think research priorities could and should change in the light of these conclusions. In the situations repre-

sented by curves B and C in Fig. 2, the small LAW means that the detection and regulation of exposure to low levels of a range of naturally occurring or synthetic carcinogens will be of limited use in extending average population lifespans⁵. A striking example is the effect of reducing tobacco smoking: although a reduction will have a significant effect on AVA, lung cancer incidence and mortality, it is unlikely to prevent cancer *per se* as a major cause of death; other types of cancer will eventually take over as causes of death. The critical issue here is therefore not the cause of death (cancer), but the time and the manner of the death at the level of the individual.

In the situation envisaged for a developing country (curve A of Fig. 2), environmental carcinogens will probably figure low on the list of health priorities, behind initiatives such as control of population growth, improvements in housing and sanitary facilities and diet, control of infectious diseases, general health care, and so on. Here, deployment of resources to control environmental carcinogens can be justified only in preventing major environmental degradation or dramatic local increases in chemically induced cancers. In fact, control of environmental carcinogens is likely to impinge on human cancer incidence only in specific situations where a point source of a potent carcinogen is encountered. The key word here is 'potent'. Even under such conditions, as with smoking, it is unlikely that the average LAW of the whole population will be affected, even though a chemically induced increase in cancer incidence may be locally observed. Examples of such a situation are the occurrence of angiosarcoma in workers exposed to vinyl chloride⁶ and the high levels of oesophageal cancer in parts of China locally contaminated with nitrosamines⁷.

Extrapolating these considerations to all environmental situations, two messages emerge. First, attention and resources should be focused on the detection of major defined sources of carcinogenic hazard: disjointed attempts to identify all potential environmental carcinogens in one or another genotoxicity test system, in our view, will ultimately be wasteful of resources. Potent carcinogens will almost always be overtly genotoxic, suggesting that primary genotoxicity tests such as the Ames *Salmonella* assay will be important in any

screening process⁸. The next and equally critical step is to assess the extent of human exposure to candidate carcinogens — both in terms of the number of people exposed and the extent and duration of exposure. Such a step could allow local increases in cancer incidence to be anticipated and prevented.

Second, given the marked differences in LAW between developed and developing countries, attempts at hazard ranking for individual chemicals at a global level are not meaningful. Carcinogenic factors of importance in one country may be of little or no importance elsewhere, so health-care resources could be unnecessarily wasted. This implies that genotoxicity test methods and biomarkers are useful only in specific situations where exposure to a suspected carcinogen is compared to the effects of other hazardous agents in the same environment, other causes of death and the local LAW(%) value. For instance, regulation of a new pesticide with a low carcinogenic potency in countries with a small LAW(%) should be related to the relative potency of the new carcinogen versus other existing (natural) ones^{9,10}. By contrast, the same carcinogen dispersed in countries with a large LAW(%) should be considered in terms of its overall influence on AVA and not just within the limited context of its rodent carcinogenicity. There is thus a need to identify the affected sub-populations as well as the carcinogenic potential of the agent of concern, and to balance these factors against concomitant hazards and the LAW(%) of the exposed population.

Finally, the more we understand the mechanisms involved in the aetiology of naturally occurring and chemically induced cancers, the more possibilities there will be for developing tools and techniques for their accurate prediction, diagnosis and early intervention. □

P. H. M. Lohman and K. Sankaranarayanan are in the MGC-Department of Radiation Genetics and Chemical Mutagenesis, State University of Leiden, 2333 AL Leiden, The Netherlands; J. Ashby is at the ICI Central Toxicology Laboratory, Alderley Park, Macclesfield SK10 4TJ, UK.

1. Böthig, S.A. *Int. J. Epidemiol.* **18**, S29–S37 (1989).
2. Fries, J. F. *Pers. Biol. Med.* **31**, 407–428 (1988).
3. Fries, J. F., Green L. W. & Levine, S. *Lancet* **1**, 481–483 (1989).
4. Harman, D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5360–5363 (1991).
5. Peto, R. In *Assessment of Risk from Low-Level Exposure to Radiation and Chemicals* (eds Woodhead, A. et al.) 3–16 (Plenum, London, 1985).
6. IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans **19**, 377–438 (1979).
7. Poirier, S., Oshima, H., de-The, G. Hubert, A. & Bartsch, H. *Int. J. Cancer* **39**, 293–296 (1987).
8. Ashby J. & Morrod, R. S. *Nature* **352**, 185 (1991).
9. Ames, B. N. & Gold, L. S. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7772–7776 (1990).
10. Ames, B. N. & Gold, L. S. *Mut. Res.* **250**, 3 (1991).

Of animals and men

SIR — I was surprised that *Nature*¹ thought it necessary to make something of a conundrum of the fact that Charles Darwin was a strong supporter of physiology and recognized the need for experiment on animals. In 1875, largely in response to a vigorous antivivisectionist campaign, the Cardwell Commission was set up. As it was drawing up its report, it became apparent that the government would respond by bringing a bill before parliament. "It was evidently desirable for physiologists to form an association which might come into communication with the minister in whose hands the conduct of the Bill would be placed. . . . Thus the antivivisection agitation was the cause of the formation of the Physiological Society. *Ex malo bonum!*"². Charles Darwin was made one of two honorary members when the society was founded in 1876. What better credentials could he have?

R. A. Chapman

School of Veterinary Science,
University of Bristol,
Park Row,
Bristol BS1 5LS, UK

1. *Nature* **356**, 644, (1992).

2. Scharpe-Schafer, E. *History of the Physiological Society during its First Fifty Years, 1876–1926* (Cambridge University Press, 1927).

SIR — Christopher Anderson (*Nature* **356**, 3; 1992) records the surprise of National Institutes of Health (NIH) scientists at the speed with which Bernardine Healy ordered an inquiry into a case of alleged abuse of animals. The surprise itself is alarming, however, in the light of the description of the conditions that inspired the complaint: ". . . 48 cats, many of which had been surgically blinded . . . a respiratory infection that many had developed during travel. The cats were inadvertently allowed to breed during quarantine, and most of the kittens died."

If complaints about such a deplorable state of affairs actually require nothing less than an "animal rights movement", then the level of ethical sensibilities within NIH would appear to be as low as the "activists" say. I found the nearly casual reference to Dr Rauschecker's ". . . investigating neural plasticity" frankly sobering, coming as it did at the end of this catalogue of horrors.

In commenting on the success that People for the Ethical Treatment of Animals enjoyed by directing a complaint to Healy, Anderson observed that it ". . . makes it likely that animal rights activists will use the tactic again". Tactic, indeed!

Daniel N. Robinson

Georgetown University,
Washington, DC 20057, USA

SIR — David Porter (*Nature* **356**, 101; 1992) leaves one crucial matter unresolved: is his ethical scoring system intended as a legal set of rules, or as a self-audited guideline for researchers?

If it were used to establish legal constraints over the use of animals in laboratories, it should also be applied to the use or keeping of animals on farms or as pets. The Royal Society for the Prevention of Cruelty to Animals would no doubt confirm that there is large-scale animal suffering outside the laboratory. Further, this is often inflicted for no better reasons than profit, neglect or just plain cruelty. Why should scientific experimentation be singled out for such detailed ethical analysis?

If, on the other hand, the system is to be used as a self-imposed guideline for researchers, surely it is unnecessary. The overwhelming majority of researchers are as concerned as Porter for the welfare of the animals they use. Attaining what would be a low (ethical) score in the proposed system would be natural to these people already. Those (very few) researchers who have a scant regard for animal suffering will simply ignore the guidelines.

The fact is that ethics are personal, and trying to draw accepted boundaries between what is and is not ethical is impossible. What really matters, in a practical sense, is what is legal. So, back to the original question: for what purpose is the scoring system intended? If it is a tool to establish legality, I believe that the scientific community should think very hard before putting it forward as a recommendation.

P. A. Leonard

Park Mansions,
London SW1, UK

■ On page 102 of Dr Porter's article referred to above, the second sentence in the second paragraph in the section 'Scoring the science' should have read: ". . . but given the fairly high probability that it will not do so, it is difficult to justify the inflicting of suffering in such inquiry". The word 'quantify' was inadvertently substituted for 'justify' in the published text. □

SIR — Frederick Goodwin, of the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA), recently created a political storm by suggesting that research on aggression in monkeys might help us to understand problems of inner-city males. Jeffrey Mervis's News article on this event (*Nature* **356**, 6; 1992) seems to me unbalanced.

It quoted individuals critical of the content as well as the language of Good-

win's statement, but it offered no contrary view. Yet officials of four societies, including the American Psychiatric Association and the Society for Neuroscience, had made public statements or written letters defending the content of Goodwin's message.

I have also been informed that there has long been tension in ADAMHA between supporters of research based on the biomedical sciences and individuals who see mental illness as mostly social in origin. Goodwin, who has been a vigorous advocate of the first view (and a most articulate opponent of animal rights extremists) has inevitably been opposed by the second group. Your reporter does not seem to have sought out this relevant background material.

There are important issues at stake for the scientific community. If we agree that Goodwin's choice of words was unfortunate (for which he promptly apologized), the deeper question is whether the message itself, defending the relevance of animal studies to problems of mental health, is to be subject to political censorship. Such a reaction seems very possible, however carefully the message is phrased. And if such animal research is discouraged, what are the implications for cancer research in mice, or for cardiac research in dogs?

Bernard D. Dayis

Harvard Medical School,
Boston, Massachusetts 02115, USA

SIR — In Donald Hayes' ranking of lexical difficulty (*Nature* **356**, 739; 1992), his most demanding test is an article from *Nature*, while his easiest is "farm workers talking to dairy cows". Simple language presumably maximizes their milk yields. I suggest a control experiment — reading *Nature* to dairy cows.

David Jones

Physical Chemistry Department,
University of Newcastle upon Tyne
Newcastle upon Tyne NE1 7RU, UK

Correction: Measuring animal well-being

In a letter from Andrew J. Wilson published on 16 April (*Nature* **356**, 556; 1992), a line was inadvertently omitted from the fourth paragraph, which should read as follows:

The major problem, however, is to decide which measurement of productivity to use. To quote from a review of this problem in respect of poultry⁵, "a change in an environment variable may reduce the number of eggs produced, but increase egg weight, leaving egg mass output the same. Depending on the measure of productivity selected, the change could be said to reduce, increase or unaffected performance." This may seem to be an extreme example, but in the course of investigating the welfare of several species of farm animal over the past few years, I have found that producers do indeed use such arguments. □

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only.

Rage and confusion hide role of HIV

Duesberg's argument that HIV is not the cause of AIDS gained no new ground at last week's alternative AIDS conference, but his message has already been taken up by a small army of followers.

Amsterdam. Dr Peter Duesberg of Berkeley, the *enfant terrible* of the aetiology of AIDS, is a little less *terrible* now. Last week, he so exaggerated his argument that human immunodeficiency virus (HIV) is not the cause of AIDS that he partly lost the sympathy even of an audience biased in his favour. Duesberg was speaking at the end of the first day (14 May) of the widely noticed alternative AIDS conference, otherwise "AIDS: A different view". The conference itself sprang from the genuine puzzlement of a group of Amsterdam physicians by recent controversies — the effectiveness of AZT in the treatment of AIDS, the diagnosis of AIDS as a contributory cause of death and, inevitably, whether HIV is a necessary and sufficient cause.

The originator of the conference is homeopathic physician Martien Brands, operating through an organization called the Foundation for Alternative AIDS Research. The meeting was organized with a help of F1100,000 from public funds. Professional qualifications were not required of the participants, who included several patients with AIDS and, sometimes, their physicians. The setting was a cavernous Protestant church converted into a conference centre, but with the pulpit, often flooded, in centre stage. Many of the speakers, as it happens, had come to preach.

Duesberg was one, but he seemed seriously to have misjudged his audience. His case is now familiar. In one of its rare published versions (*Proc. natn. Acad. Sci. U.S.A.* **88**, 1575; 1991), he argues from the cryptic character of HIV (it can be recovered from only a small proportion of CD4⁺ cells in infected persons, while the disease appears often many years after viral antibodies) and the quirks of epidemiology (people with haemophilia and HIV are less susceptible to AIDS than infected male homosexuals, for example) to conclude that something other than HIV must be the cause of AIDS.

Duesberg believes that the taking of what are called 'recreational' drugs, and even of therapeutic drugs, can and does trigger AIDS; the evidence is circumstantial; drug-taking in the broad sense has increased with AIDS.

Last week's presentation of this case was a slimmed-down version for a mixed audience; little about cells, more about epidemiology. But at the outset, the audience was plainly ready for anything. The instant that Duesberg had finished arranging his slides in their carousel, the hall was echoing to the noise of hands clapping in just the manner that theatre-goers associate with what they

call a *claque*. Some members of the audience let out occasional whooping sounds, like those from people watching ski-races.

Duesberg is a neatly built and restless man, more comfortable pacing up and down than standing at a podium. For his presentation last week, he wore a cerise shirt with the sleeves neatly folded back half-way up the forearms. He did not pull his punches, but rather chose his words for their dramatic effect.

The message was simple: "HIV does not cause AIDS." "In 1984, Margaret Heckler [then the US Secretary of Health and Human Services] stood up alongside [Robert] Gallo and announced that they would have a vaccine against AIDS in three or four years. But see what has happened! The HIV hypothesis has so far failed to save a single life."

The argument? First, a table showing the incidence of HIV infection and of AIDS in different groups in different parts of the world. Pointing to acknowledged differences in the rates at which different groups (male homosexuals, female heterosexuals and haemophiliacs) with HIV convert to AIDS, Duesberg mocked the efforts of clinicians to make sense of the figures. "They cannot even tell whether a particular patient will go down with diarrhoea or Kaposi's sarcoma (KS)."

Emphasizing the apparent differences between the data for Africa and the rest of the world, Duesberg said that "in the United States, we convert 3 per cent of those with HIV into AIDS patients", but in Africa, the ratio is "only 0.3 per cent". He also stressed the differences of the sex ratio of those carrying HIV (near-equality in Africa, a predominance of males in the United States) and claimed that the adventitious infections that characterize AIDS in the United States are absent in Africa. "AIDS is at least two different diseases." Later, he declared that "to call AIDS a single disease is entirely unproductive". His view seems to be that there are as many diseases as there are causes of death from AIDS — Kaposi's sarcoma and the like.

So where do drugs come in? This part of the argument is also circumstantial. Duesberg noted that drugs of all kinds are recent social innovations. He sustained the assertion that recreational drugs are the driving force behind the growing prevalence of AIDS with statistics showing the rising numbers of arrests and seizures of drugs by the US Customs. Nobody raised the question whether the US Customs' success in recent years may be an index of the attention paid

to the drugs problem in the United States.

Duesberg's rhetorical style is mocking of the adherents of the 'HIV hypothesis'. Thus he made fun of the definition of AIDS promulgated by the US Centers for Disease Control (CDC) with a series of equations. "If you have TB and HIV, then you have AIDS. In other words, $TB + HIV = AIDS$ ". "So", he went on, $TB - HIV = TB$. "HIV, in other words, is just a way of stigmatizing some sick people as people who have AIDS. That produced a burst of whooping from an otherwise subdued audience. But one of Duesberg's mistakes may have been the fun he poked at the adherents of the 'HIV hypothesis'. Few people, and certainly not those diagnosed, even stigmatized, as suffering from AIDS are easy with the notion that there is polemical fun to be won from the dilemma in which they find themselves. The sight of Duesberg enjoying himself so hugely at the expense of lesser mortals who have managed so grossly to misunderstand AIDS was not endearing.

Another was his failure, even at the end of a long day, to respond to the counter-arguments flagged prominently in advance. One of the most cogent of these was that due to Professor R. Coutinho, an epidemiologist from the Municipal Health Centre at Amsterdam, who courageously (in the setting) rounded off his presentation with the declaration (on a slide) that "HIV is the cause of AIDS". So far as Duesberg was concerned, Coutinho might never have spoken.

Similarly, Duesberg seems not to have grasped a point made in discussion by Dr Luc Montagnier from the Pasteur Institute in Paris that the quality of the statistics for Africa was hardly such as to sustain the conclusions drawn from them, although Duesberg was distinctly heard to use the phrase "in that country" to refer to "Africa", which may simply be a measure of the distance between Berkeley and the African continent.

Of course, all these issues might have been dealt with if Duesberg had been able to complete his presentation. But 50 minutes into his 30-minute allocation, Duesberg sensed the chairman breathing down his neck. By sleight of hand, he managed to project three or four extra slides while fending off the call to dinner with a hurried commentary at about 300 words a minute, but finally left the last slide showing and sat down in a sulk. After a few seconds, there followed thunderous applause and more whooping.

Worse was to come. In discussion, the

Ghanaian physician responsible for Uganda's AIDS programme explained that his dilemma was to know what to tell his colleagues when he returned from Amsterdam — "There are so many lives at risk". He then recited from a World Health Organisation document figures for the annual incidence of new African AIDS cases, which totalled 19,500-odd in 1989 — just a little less than the figure from which Duesberg had derived 0.30 per cent as the ratio of AIDS to HIV infections. There was relief all round. But then it emerged that the quoted figures were for Uganda alone. What they imply for the rest of the country called Africa was left to the imagination.

That contretemps illustrates one obvious weakness of meetings on technical matters where the interests of the audience are much more general. Often (but not always), people referred to published research, but did not supply the references on the spot, no doubt for fear of making their presentations tedious. Attempts at technical discussion were invariably diluted by protests against the 'HIV hypothesis' by or on behalf of people with AIDS. And clashes between Duesberg and other specialists (notably Coutinho) during panel discussions often boiled down to assertions such as "You're wrong!".

Disappointment among Duesberg's followers with his performance last week was muted, but real. Brands said he wished that Duesberg had been more measured. An open supporter said that Duesberg had "gone over the top". Another was offended by the jokes, directed though they were at the malign originators of the HIV hypothesis. One of those, Montagnier, came as near as such an equable character can to the state of being known as "hopping mad". He said that he had agreed in good faith to come to this conference, but would not attend another.

That will be a pity, for last week's meeting brought into the open the AIDS community's deep anger at the 'HIV hypothesis' and the use of drugs such as AZT in treatment of the disease as well as its determination that something else should happen.

But why should the notion that HIV is the cause of AIDS engender such anger? There is no mystery. The obvious explanation is simple. The HIV hypothesis is often presented to patients by their physicians as a death sentence. For many, it is literally that. Who can blame those given such bad news for harbouring rage, sometimes uncontrolled?

The opening speaker last week, Michael L. Callen from New York, was diagnosed as having AIDS almost exactly ten years ago. He attributes his survival to his physician's prescription of drugs to guard against the adventitious infections and other conditions from which AIDS patients are at risk. He held up a plastic bag half-filled with capsules, saying that he takes 56 of them each day.

Callen is not for the avoidance of "Western medicine", but complains that researchers have paid too little attention to the reasons why he and others have survived for such long periods. His own impression (based on 48 other long-term survivors) is that there is a distinct "survivor personality" (feisty, even aggressive, and knowledgeable). "Lifestyle changes" are crucial, but people who take AZT (in Callen's opinion) "are often worse off than if they did nothing at all".

While acknowledging that there is some support for the HIV hypothesis, he argues that all plausible alternatives should be investigated, "and receive funding". Even allopathic medicine needs debate, despite the criticism that it merely makes people feel better. "Making people feel better is not to be discounted where AIDS is concerned."

This point was put more generally by Dr H. Albonico, a physician from Langau in Switzerland, "The AIDS dogma", he said, "is one of the elements of the pathogenesis of the disease: fear kills." It would obviously be preferable if physicians could tell people with HIV that "the solution may be in their own minds". Whence last week's discussion of psychoimmunology or psychoneuro-immunology. Dr R. Root-Bernstein had earlier given a well-documented account of how circulating immune complexes may be responsible for some of the symptoms of AIDS, dementia in particular. But the standard presentation of how the hypo-thalamus may influence the immune system and thus enable a person to influence his state of health by thought, innocent (as it usually is) of documentation, quickly turned into a kind of consensus that the question "Can you use your mind to control your disease?" should be answered in the affirmative.

The several homeopathic physicians at the meeting were mostly genuinely concerned. They have a problem. If AIDS is a novel disease, what assurance can there be that traditional homeopathic remedies will serve any purpose? Dr M. Strange from the Lavender Hill Homeopathic Centre in London said that the more time AIDS patients "spend in the clinic talking about the results of their blood tests", the worse off they will be.

The principal medicine in the field at present, AZT, is of course at odds with homeopathic medicine, but that does not adequately explain why this drug and its manufacturers are the targets of such open hostility. The most implacable of the critics was J. Lauritsen, a market researcher turned AIDS journalist, who claimed to have found such serious errors in four key articles reporting the results of clinical trials with AZT that he alleged that the studies were "fraudulent". The estimated 100,000 people now being treated with AZT are the victims, Lauritsen said, of "pharmacogenic manslaughter".

The same message was repeated by several speakers, both in formal presentations

and in questions and comments from the floor. The roots of this anger against AZT, even from the point of view of what Lauritsen described as the "official paradigm", deserve to be better understood. They are that to accept that AZT may be beneficial is to accept the "HIV hypothesis", that the drug is only a palliative and not a cure, that it cuts across the wish that AIDS might be abated by taking sufficient thought (and a healthy diet) and that the side-effects are often unpleasant).

The other side of that coin is the concern of AIDS organizations with good health, represented last week by the British organization "Positively Healthy" and by the New York organization HEAL. In the view of both, and of other speakers last week, the best treatment for AIDS at present is to eat well, to avoid alcohol and other drugs and to be as cheerful as possible.

Positively Healthy, in a list of dos and don'ts for people with AIDS, says (*inter alia*, "Learn to question everything you have been told . . . It is mostly lies." The message is no doubt a reinforcement of a resolve to live well and sensibly. Yet both organizations claim they cannot get the funds for the research to prove their point — HEAL, for example, says that it is easier to raise \$50,000 for a course of treatment with AZT than for a research project to study the effectiveness of its regimen. One would have thought that if such a study could be mounted, the outcome would be not just enlightenment, but some abatement of corrosive anger.

For the rest, last week threw up little in the way of science, although Dr Joan McKenna from TBM Associates in Berkeley, California, argued interestingly that syphilis may be the underlying cause of the immune deficiency of AIDS, with HIV a kind of irrelevant accompaniment. Certainly there are striking similarities in the behaviour of the immune system in syphilis and AIDS. Unfortunately the argument (which has not been published) rests on an untested notion of the thermodynamics of living things.

Or is syphilis just another co-factor? Last week's meeting was full of talk of co-factors. Montagnier, describing his reasons for believing that mycoplasmas are involved, said clearly what he meant: HIV is a necessary but not, without the co-factor, a sufficient cause of AIDS.

Not everybody was as Cartesian. Age, diet, 'lifestyle' and even the effects of the "medicine of fear" on the immune system were loosely labelled "co-factors".

Brands hopes to present a report on last week's meeting to the annual international conference on AIDS next month, also in Amsterdam. It will be interesting to learn what he says. His hope at the outset of last week's meeting — "Let the unheard voices of science express themselves" — may have been partly fulfilled. The need now is that they should use language others understand.

John Maddox

Yeast branches out

Ira Herskowitz

As we learn more and more about yeasts at a molecular level, we discover more and more ways in which their cells are like those of metazoans. Yeasts have homeodomain proteins that programme cell type¹, heterotrimeric G proteins involved in signalling² and *cdc2/CDC28* protein kinases governing the cell cycle³. Because yeasts are single-celled fungi they might be expected to have particularly close functional and evolutionary relationships to other fungi, and revelations emerging from several laboratories show that this is indeed the case.

First, Gimeno *et al.*⁴ report that the budding yeast *Saccharomyces cerevisiae*

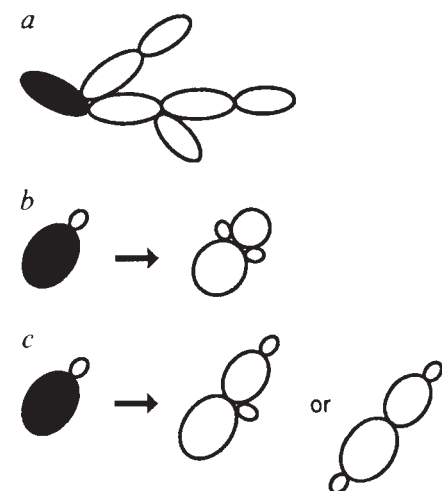


FIG. 1 Different morphological forms of budding yeast. *a*, Pseudohyphal mode of growth, which is exhibited by *a/a* diploid cells when starved; the initial cell is in black. (From ref. 4.) *b*, Axial pattern of bud formation, exhibited by *a* or *α* haploid cells; and *c*, bipolar pattern of bud formation, exhibited by *a/a* diploid cells. (From ref. 10.)

can form strings of elongated cells resembling the hyphae of filamentous fungi; yeast is thus related morphologically to *Neurospora* and *Aspergillus* (which like yeast are ascomycetes but are filamentous). At the same time, other groups⁵⁻⁸ have shown that several higher fungi — basidiomycetes such as the plant pathogen *Ustilago maydis* ('corn smut') and the wood-rotting fungus *Schizophyllum commune* — contain regulatory proteins governing their life cycles that are like the well-studied *a1-a2* repressor of yeast⁹.

Saccharomyces cerevisiae is a budding yeast, producing daughter cells that are separate from the mother cell. Hyphal growth, in contrast, yields chains of connected cells. Gimeno *et al.*⁴ have now observed that, when starved of nitrogen, yeast cells exhibit a mode of growth that

is reminiscent of filamentous growth. Under these circumstances, yeast produces a chain of cells that radiates out from the centre of a colony: the newly born cells bud away from their mother and become strikingly elongated, changing from an egg shape to a sausage shape (Fig. 1*a*). Filamentous fungi produce cells that remain connected to one another, yielding branched chains or hyphae. The cells of the pseudohyphae of budding yeast, by contrast, can be separated; they are also able to grow into the surrounding agar, possibly because they produce degradative enzymes that allow them to invade this new 'habitat'. These degradative capacities are reminiscent of the versatile properties of other fungi which can break down materials such as wood, fruit and shower curtains.

Some aspects of the pseudohyphal mode of growth appear to be an elaboration of differences in budding pattern. Well-fed yeast cells exhibit two natural forms of morphogenesis: axial, characteristic of haploid cells (Fig. 1*b*), and bipolar, characteristic of diploid cells (Fig. 1*c*)^{10,11}. These budding patterns, like haploidy and diploidy themselves, can be explained as having certain adaptive advantages. That of diploid cells is like the pattern seen in the pseudohyphal mode of growth, and presumably allows yeast strains to grow away from each other and thus lessens competition for nutrients. The adaptive value of the budding pattern of haploid cells can be contrasted by the proposal¹² that it optimizes mating between *a* and *α* cells in homothallic strains. In these strains, haploid cells (of mating type *a* or *α*) give rise to cells of the other mating type in a precise pattern (Fig. 2)¹³, so that *a* and *α* cells are perfectly positioned to mate with one another. Mating yields diploid cells (the *a/a* cell type) which are then programmed to grow away from each other.

The critical feature of diploid yeast cells, responsible for their many distinctive cell specializations, is their possession of the repressor *a1-a2*. This protein is formed by combining the *a1* polypeptide encoded by the *a* mating type locus with the *a2* polypeptide encoded by the *α* mating type locus (refs 14, 15, and C. Goutte and A. D. Johnson, personal communication). From sequence considerations, it has been argued that both *a1* and *a2* are homeodomain proteins¹⁶; the three-dimensional structure¹ of *a2* shows that *a2* clearly is. The role of *a1-a2* in conferring properties of the diploid cell

is extensive: it represses transcription of genes necessary for mating and for mating type switching⁹ and it turns off synthesis of a repressor of meiosis and spore formation¹⁷. *a1-a2* is also thought to be responsible for the differences between the budding pattern of haploid and (*a/a*) diploid yeasts, in this case by inhibiting the proteins (*BUD3* or *BUD4*) necessary for axial budding¹¹.

What of the other part of this story, that bringing in the basidiomycetes? Here the news is the discovery of proteins analogous to the yeast *a1-a2* that regulate basidiomycete life cycles — such proteins have now been reported in *Ustilago maydis*⁵, *Schizophyllum commune*^{6,7} and *Coprinus cinereus*⁸ (Fig. 3). As described in News and Views earlier this month¹⁸ and by Banuett¹⁹, in *U. maydis* a heteromultimeric species is formed between the *b-E* polypeptide encoded by one allele of the *b* locus and *b-W* polypeptide encoded by another allele of the *b* locus⁵. There are additional morphogenetic parallels between *U. maydis* and budding yeast. *Ustilago* has two morphological forms, a haploid form that is yeast-like (and which grows by budding) and a dikaryotic form produced by mating of haploids that is filamentous¹⁹. As in budding yeast, where the pseudohyphal form requires cells to contain *a1-a2*, production of the filamentous form by *Ustilago* requires cells to contain, for example, the *W1-E2* (or *W2-E1*) product of the *b* locus⁵. It may be that the products regulated

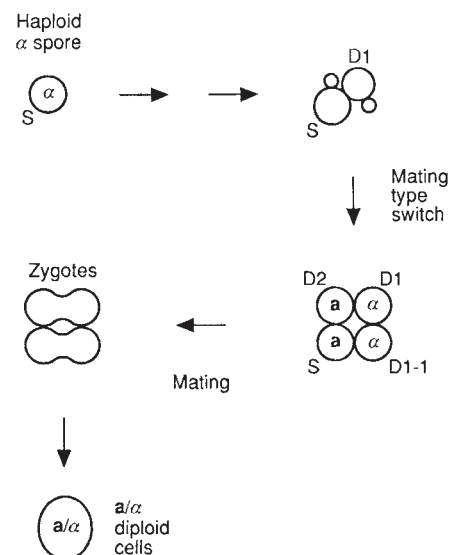


FIG. 2 Formation of diploids by mating between siblings in homothallic yeast strains. The switching of yeast cells from *α* to *a* occurs in a defined pattern in a mitotic cell lineage^{9,13}. The axial budding pattern is thought to facilitate formation of diploids by juxtaposition of *a* and *α* cells¹². *S* is a spore cell; *D1* is the first daughter of *S*. Mating type switching occurs in cells that carry the *HO* gene; *HO* is repressed in *a/a* cells by *a1-a2*.

by W1-E2 and $\alpha 1$ - $\alpha 2$ that are responsible for morphogenetic differences in these two organisms will turn out to be similar.

Thus budding yeast exhibits some of the properties of the filamentous fungi and they in turn share some of the gene-regulatory mechanisms of yeast that control these properties. It is important, however, not to overstate the similarities between budding yeasts and their fungal brethren: the filamentous fungi present many distinctive opportunities for studying cell-biological processes

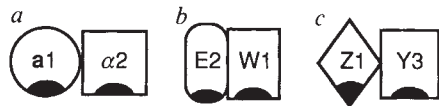


FIG. 3 Heteromultimeric regulatory proteins are molecular signatures for specialized cells of yeast and other fungi. *a*, In *Saccharomyces cerevisiae* *MATa* codes for $\alpha 1$, and *MAT α* codes for $\alpha 2$, which form the heterodimeric repressor $\alpha 1$ - $\alpha 2$. *b*, In *Ustilago maydis* *b1* codes for E1 and W1, and *b2* codes for E2 and W2, which can form two different types of heteromultimers (one of which is shown). *c*, In *Schizophyllum commune* *Aa1* codes for Z1 (and several other polypeptides), and *Aa3* codes for Y3 (and several other polypeptides), which can form several different types of heteromultimers (one is shown). The black segments of the subunits indicate known or presumptive homeodomains.

(nuclear migration and morphogenesis for instance), and degradative processes and other biotransformations, and they are vitally important plant and animal pathogens. We can look forward to further revelations about the similarities and differences between yeast and its fungal brethren. □

Ira Herskowitz is in the Department of Biochemistry and Biophysics, School of Medicine, University of California San Francisco, San Francisco, California 94143-0448, USA.

Geminga joins the class

Charles D. Bailyn

THE undercurrent of elation in Halpern and Holt's paper on page 222 of this issue¹ on the 'Geminga' astronomical γ -ray source is unmistakable. The authors "consider the mystery of Geminga largely solved", laying to rest a puzzle of two decades standing, with important implications for understanding high-energy sources in our Galaxy in general. The sentiment is justified for, at Halpern and Holt's suggestion, Bertsch *et al.* re-examined data from the source from the satellite-borne Compton Gamma Ray Observatory and confirmed the conclusion. Their paper², and further confirmation by Bignami and Caraveo³, will appear in the next issue of *Nature*.

The γ -ray sky is dominated by diffuse emission and three bright point sources. Two of these, the famous pulsars associated with the Crab and Vela supernova remnants, as soon as they were discovered were identified with objects prominent in other wavelengths, and have become among the most intensively studied of astronomical objects. The third source, however, appeared to have no radio, X-ray or optical counterpart, and was therefore dubbed Geminga, a pun on the constellation of Gemini in which it is located and an Italian expression meaning "it's not there"⁴.

Indeed it would be more than 10 years before an unusual faint X-ray source was proposed as a counterpart for Geminga⁵, and another five before the discovery of a faint optical object in the X-ray error box⁶. Now, finally, the identification of the X-ray and optical source with Geminga has been unambiguously confirmed. Furthermore, the source itself is shown to be of the same general class as the Crab and Vela pulsars. The similarities and differences among the three sources will provide interesting constraints on pulsar emission mechanisms.

The key which has unlocked Geminga's secrets is the discovery by Halpern and Holt of pulsations with a period of 0.237 seconds in X-rays. The authors used a long exposure with the Rosat X-ray observatory of the purported X-ray counterpart to reveal the pulsations. It had not been possible to search for such short pulsations in γ -ray data, because the paucity of γ -rays leads to spurious periodicities on sub-second timescales. But with the Rosat periodicity in hand, Bertsch *et al.*² quickly found the same pulsations in data from the currently operating Compton Gamma-Ray Observatory; and Bignami and Caraveo successfully searched archival γ -ray data from the 1970s, exclaiming that "it was there all the time"³. The

identical pulsation periods in X- and γ -rays confirm that Geminga and the X-ray source are one and the same.

A pulsation period of this length strongly suggests that Geminga is a young pulsar, in which the pulsation period is associated with the rotation period of a highly magnetized neutron star. Geminga is not observed in the radio, as are the Crab, Vela and many other pulsars, apparently because its magnetic poles do not point sufficiently closely in our direction. Such radio quiet pulsars have long been inferred to exist; what makes Geminga unusual is that it is so nearby that its γ -ray emission is easily seen. Indeed, the γ -ray data indicate that Geminga may be but 30 parsecs away from us (only 25 times further than the nearest star), making it the closest identified neutron star.

The presence of the pulsation signal in data two decades old allows Bignami and Caraveo to confirm the observation of Bertsch *et al.* that the pulse period is slowly but steadily lengthening. Such behaviour is observed in all radio pulsars, and is generally accepted to be due to a braking force exerted by the strong magnetic field on the neutron star's rotation. The inferred magnetic field for Geminga is right in the middle of the range observed for young radio pulsars, which clinches the case for the view that Geminga is a member of this class.

Now that the general nature of Geminga has been established, much work remains to be done to knit this astronomical curiosity into the growing fabric of high-energy astrophysics. Theories of pulsar emission mechanisms will have a new constraint to satisfy; indeed Bertsch *et al.* suggest that one popular model may have to be discarded. More sensitive searches for Geminga-type objects with the new generation of orbiting high-energy observatories may improve the statistics of radio pulsars in various ways. But the mere fact that this mysterious object can now be fitted into current ideas about high-energy astrophysics is in itself enough to delight and invigorate those working in this rapidly advancing field. □

Charles D. Bailyn is in the Department of Astronomy, Yale University, PO Box 6666, New Haven, Connecticut 06511, USA.

1. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. *Cell* **67**, 517-528 (1991).
2. Marsh, L., Neiman, A. M. & Herskowitz, I. A. *Rev. Cell Biol.* **7**, 699-728 (1991).
3. Nurse, P. *Nature* **344**, 503-508 (1990).
4. Gimeno, C. J., Ljungdahl, P. O., Styles, C. A. & Fink, G. R. *Cell* **68**, 1077-1090 (1992).
5. Gillissen, B. *et al.* *Cell* **68**, 647-657 (1992).
6. Stankis, M. M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
7. Specht, C. A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
8. Kües, U. *et al.* *Genes Dev.* **6**, 568-577 (1992).
9. Herskowitz, I. *Nature* **342**, 749-757 (1989).
10. Hartwell, L. *Nature* **352**, 663-664 (1991).
11. Chant, J. & Herskowitz, I. *Cell* **65**, 1203-1212 (1991).
12. Nasmyth, K. A. *Rev. Genet.* **16**, 439-500 (1982).
13. Strathern, J. N. & Herskowitz, I. *Cell* **17**, 371-381 (1979).
14. Goutte, C. & Johnson, A. D. *Cell* **52**, 875-882 (1988).
15. Dranginis, A. M. *Nature* **347**, 682-685 (1990).
16. Scott, M. P., Tamkun, J. W. & Hartzell III, G. W. *Biochim. biophys. Acta* **989**, 25-48 (1989).
17. Covitz, P. A., Herskowitz, I. & Mitchell, A. P. *Genes Dev.* **5**, 1982-1989 (1991).
18. Egel, R. *Nature* **357**, 23-24 (1992).
19. Banuett, F. *Trends Genet.* **8**, 174-180 (1992).

1. Halpern, J. P. & Holt, S. S. *Nature* **357**, 222-224 (1992).
2. Bertsch, D. L. *et al.* *Nature* (in the press).
3. Bignami, G. F. & Caraveo, P. A. *Nature* (in the press).
4. Bignami, G. F., Caraveo, P. A. & Lamb, R. C. *Astrophys. J.* **272**, L9-L13 (1983).
5. Halpern, J. P. & Tytler, D. *Astrophys. J.* **330**, 201-217 (1988).

In the air

ASTUTE seismologists may be able to hear distant volcanic eruptions, H. Kanamori and J. Mori have discovered (*Geophys. Res. Lett.* **19**, 721–724; 1992). The authors suggest that atmospheric waves set up by the eruption at Pinatubo were responsible for a puzzling, two-hour seismic rumble recorded worldwide on 15 June last year. The signal was weak, but the authors identified a strong component with a period of 228 seconds — far too long to be associated with any activity in the volcano's magma chamber. But there are atmospheric oscillations with natural periods around 250 seconds which could have been set up by heat from the eruption. Indeed, a contemporary satellite image of Pinatubo and its environs shows a temperature anomaly extending over 300 km from the volcano, completing the authors' argument. The result was a remarkable cycle of excitation — from land to atmosphere and back to land.

Vampire wedding

DURING development of the *Drosophila* retina, formation of the R7 photoreceptor neuron is induced by contact with its neighbour R8. The process is mediated by the interaction of the cell-surface tyrosine kinase *sevenless* (*sev*), expressed on all the photoreceptors, and its R8-specific ligand *bride of sevenless* (*boss*). R. L. Cagan *et al.* (*Cell* **69**, 393–399; 1992) have used antibodies to different regions of *boss* to examine its fate after contact with *sev*. Remarkably, although *boss* is thought to contain no fewer than seven membrane-crossing helices, they find that the entire protein is taken up together with *sev* by the nascent R7 cell. The authors speculate that this may occur by receptor-mediated endocytosis of small patches of R8 membrane, although they cannot rule out other possibilities.

Enigma variations

COMPUTER reconstructions of a flawless Sphinx of Giza preside over a paper by M. Lehner in the *Cambridge Archaeological Journal* (**2**, 3–26; 1992). The monument was carved from solid limestone 4,600 years ago. It then fell prey to robbers and the ravages of shifting sands, until restoration and further adornment in the 18th dynasty 1,200 years later. Millennia of neglect then ensued. The images of the restored, 18th dynasty Sphinx — complete with beard, cobra symbol and chest statue — are the upshot of a four-year programme which resulted in a full set of contoured, scale drawings of the monument, then transformed into three-dimensions by computer graphics. The reborn Sphinx will no doubt be appearing on theme-park screens before too long.

In the pink (and yellow)

Jeremy J. D. Greenwood

'SMALL science' is alive and well, as a paper on the land snail *Cepaea hortensis* in the latest issue of *Proceedings of the Royal Society*¹ shows. Armed with little more than polythene bags, notebook and calculator, Robert Cameron has made a notable contribution to a long-standing debate in evolutionary genetics.

Cepaea nemoralis and *C. hortensis* are classic subjects of study. They display genetic polymorphisms for both the colour and the banding patterns of their shells, and there has been intense debate over the relative importance of natural selection and of random genetic drift in maintaining the polymorphism and in determining the morph-frequency differences between populations. As so often, a problem has been that selective forces may be powerful enough to override random events and yet have such subtle effects that they cannot be detected. But by use of a particularly well-focused sampling programme, Cameron has been able to establish morph-frequency changes of only a few per cent and to show that the changes are almost certainly associated with selection.

In cloudless weather, the ground in open countryside cools rapidly after sunset because of radiative heat loss to the sky. Cool air near the ground surface flows downslope and may accumulate in valley bottoms, which may as a result become many degrees cooler than the nearby valley sides. A quarter of a century ago, Cameron surveyed *C. hortensis* populations in a 10 × 10 km area in south-east England, in a region of chalk hills with such strongly incised topography that many of the valley bottoms are notorious frost hollows. There he found that the snail populations of valley bottoms tended to have higher frequencies of yellow shells than those on valley sides. This, and similar evidence from elsewhere, fits in with experimental evidence that snails with yellow shells are more resistant to cold than are those with pink shells^{2–4}. Broad geographical patterns of the distribution of yellows also fit in with this conclusion^{2–4}.

Cameron's original study took place shortly after a particularly cold, anticyclonic winter, when selection favouring yellow shells in valley bottoms must have been particularly intense. Since then, winters have been milder and Cameron surmised that the selection favouring yellow shells in valley bottoms should have been generally weaker. He therefore returned to the area and re-sampled as many of the *C. hortensis* populations as possible, carefully return-

ing to within a few metres of his original sampling sites. A number of populations had become extinct because of road-building and similar activities. In some places morph-frequencies had altered so much that the changes were statistically significant even at a single site. Such large changes were always associated with marked alterations in the habitat.

Because his original survey had shown that morph-frequencies were associated not only with topography but also with habitat (in a way consistent with the classic effects of selection by visual predators⁵), Cameron excluded these sites from further analyses. At the remaining sites, there was a clear pattern: although morph-frequencies had not changed significantly at any of the individual sites, there was a significantly consistent trend for the frequency of yellows to have dropped in valley-bottom sites between the two surveys but no general trend in sites on valley sides. This accords exactly with expectations based on the hypothesis that the excess of yellows in the valley-bottom sites in 1964–66 was largely a result of the unusually severe winter of 1962–63, because it is in valley bottoms that cold air settles. The subsequent milder winters have not, on average, favoured yellows so strongly in valley bottoms.

The gene-frequency change detected by Cameron is, by normal standards, minute: the mean frequency of the yellow allele in the area was 87.7 per cent in 1964–66 and 85.9 per cent in 1990. That he was able not only to detect it but also convincingly to ascribe it to a particular cause is because he concentrated on individual sites, not only relocating his original sample sites carefully but also taking into account that sites differed in the extent to which they would be expected to show morph-frequency changes. The study demonstrates the over-riding value in ecology of a sound knowledge of the organism on which one works, of having a clear focus to one's investigation, and (not least) of long-term thinking and meticulous fieldwork. □

Jeremy J. D. Greenwood is at the British Trust for Ornithology, Thetford, Norfolk IP24 2PU, UK.

1. Cameron, R. A. D. *Proc. R. Soc. B* **248**, 181–187 (1992).
2. Jones, J. S., Leith, B. H. & Rawlings, P. A. *Rev. Ecol. Syst.* **8**, 109–143 (1977).
3. Clarke, B., Arthur, W., Horsley, D. T. & Parkin, D. T. in *Pulmonates Vol. 2A* (eds Fretter, V. & Peake, J. F.) 219–270 (Academic, New York, 1978).
4. Cain, A. J. in *The Mollusca Vol. 6* (ed. Russell-Hunter, W. D.) 597–647 (Academic, London, 1983).
5. Cain, A. J. & Sheppard, P. M. *Genetics* **39**, 89–116 (1954).

Flying the first class flag

Peter Parham

In rats, mice and humans, two genes related to the multidrug-resistance family of transmembrane transporters lie within the class II region of the major histocompatibility complex (MHC)¹⁻⁴. They are thought to encode a pump which delivers peptides to class I MHC molecules in the endoplasmic reticulum, and their discovery in late 1990 initiated one of immunology's periodic races. Since then, the back half of *Nature* has seen a stream of papers describing the genes¹⁻³ and their rescue of defective class I MHC function in mutant cells⁴⁻⁹. Meanwhile, the front half has had, in the words of one contributor¹⁰, "a great deal of speculation about their involvement in antigen processing and peptide binding to class I major histocompatibility complex molecules"¹⁰⁻¹³.

The evidence implicating these genes in peptide transport is circumstantial but substantial, and on page 211 of this issue¹⁴ Powis *et al.* add some more; they show that polymorphism of a rat 'transporter' gene determines the peptides loaded by a class I molecule and thus its presentation of antigens. Although it can still be argued that the effect of the transporter-like protein is to select peptides through some function other than transmembrane peptide transport, such possibilities recede. Irrespective of the protein's precise function, the complexity of its polymorphism reveals the potential for these genes to influence T-cell responses, disease susceptibility and transplant rejection.

The story began with the discovery¹⁵ that alloreactive T-cell recognition of the rat class I molecule RT1.A^a was dramatically affected by another MHC locus, the class I modifier (*cim*). Of two alleles defined, *cim*^a dominates *cim*^b in heterozygous antigen-presenting cells. Kinetic differences in the maturation of RT1.A^a also correlated with *cim* type, homozygosity of the *b* allele causing retention within "the endoplasmic reticulum or cis-Golgi for an abnormally long time". Such effects were detected by monitoring glycosylation and for some time it was thought *cim* might encode a glycosylating enzyme¹⁶.

Mapping *cim* to the class II region¹⁷ and scouring appropriate

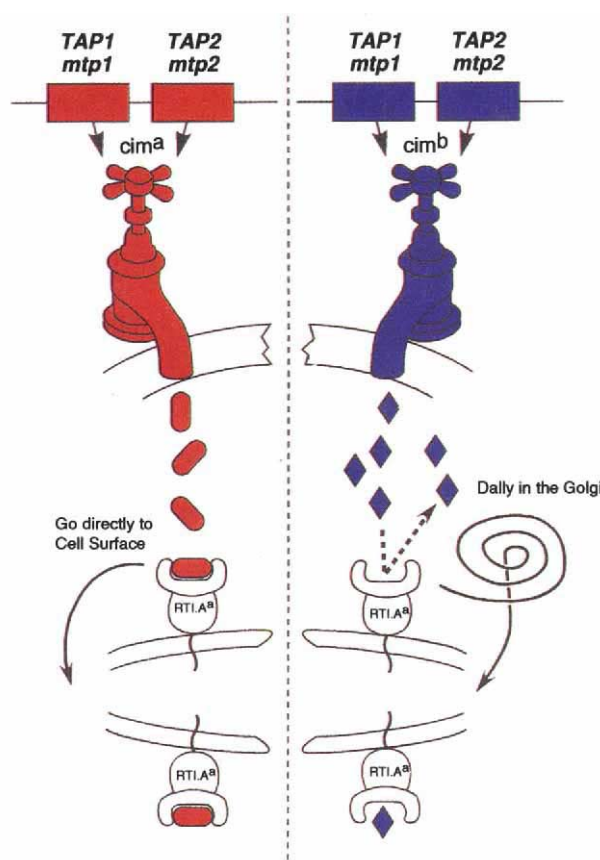
cosmids produced the putative transporter genes, *mtp1* and *mtp2* (see figure). Powis *et al.* now show that *cim* and *mtp2* are one and the same. They have sequenced *mtp2* from two *cim*^a and two *cim*^b rat strains. Although all four alleles encode proteins of different amino-acid sequence, they clearly divide along *cim* lines. Whereas sequences associated with the same *cim* type differ by just 1-3 residues, 25 amino-acid substitutions distinguish *cim*^a from *cim*^b. Furthermore, almost all of these differences are in the amino-terminal membrane-spanning domain, which is hypothesized to form the peptide channel and in which a single substitution alters the specificity of other

transporters. Proof that *cim* and *mtp2* are identical is that transfection of an *mtp2* allele confers *cim* type, both in terms of T-cell recognition and the rate of intracellular RT1.A^a transport. Moreover, the chromatographic profiles of peptides co-purifying with RT1.A^a molecules from cells of *cim*^a and *cim*^b type are distinct, showing the heterodimeric protein formed by the *mtp1* and *mtp2* genes directly affects the selection of peptides bound by the class I molecule.

The unexpected complexity of differences between *mtp2* alleles is comparable to that seen for 'highly polymorphic' class I MHC genes. Furthermore, as Powis *et al.* show, *mtp2* polymorphism can have as profound an effect upon class I antigen presentation as a change in the class I molecule itself. Heterozygosity at class I genes is believed to confer advantage to the individual by increasing the range of presentable peptides¹⁸. Similar arguments can be applied to transporter function — heterozygous expression of transporters with different peptide specificity makes a wider range of peptides available to class I molecules in the endoplasmic reticulum. Benefits to be had from such an arrangement would increase with the difference in peptide specificity, favouring evolution of increasingly divergent alleles and loss of intermediate forms, as seems to have happened in rats.

Implicit in the story told by Powis *et al.*¹⁴ is that linked transporter and class I alleles can co-evolve compatible peptide specificities. Normally the RT1.A^a molecule is linked to *cim*^a, the allele by which it is efficiently loaded (see figure). Indeed, in heterozygous rats RT1.A^a seems to be loaded exclusively by *cim*^a as the effects of *cim*^b on the molecule are undetectable. Only within the laboratory, in recombinant animals and transfected cells, is RT1.A^a forced to do business with *cim*^b. The ten rat MHC haplotypes examined so far divide equally between the *cim*^a and *cim*^b types¹⁷. If, in general, rat class I alleles have co-evolved with their linked transporters, one prediction is that evidence for these alliances will exist in the sequences of the RT1.A alleles, most of which remain uncharacterized. A second is for recombination in the RT1.A to *mtp2* region to be suppressed between haplotypes of different *cim* type.

In pondering the generality of



Turning on the peptide taps, a representation of the different peptide supply to the rat class I molecule RT1.A^a in *cim*^a (left) and *cim*^b cells (right). On discovery of the human transporters, Trowsdale *et al.*² found they were 'really interesting new genes', dubbing them *RING4* and *RING11*. Though still interesting, they are no longer new and given the competing nomenclatures from other groups³ they have been renamed. The WHO committee for factors of the HLA system named the human homologues of *mtp1* and *mtp2*, *TAP1* and *TAP2* respectively. (*TAP1* replaces *RING4*, *Y3* and *PSF1*; *TAP2* replaces *RING11*, *Y1* and *PSF2*). The acronym, *TAP*, is thought to be sufficiently flexible that no matter what the precise function of the protein turns out to be, an appropriate name can be conjured from the three letters. For the moment it is 'transporter associated with antigen processing'²⁷.

their findings, Powis *et al.* draw attention to a now famous family in which antigen presentation and peptide loading of the human class I molecule HLA-B27 is unusual and where it is suggested a transporter polymorphism *à la cim* is at work¹⁹. On the other hand, the search for polymorphisms in the human transporter genes, *TAP1* and *TAP2* (see figure), has yet to reveal anything approaching the drama seen in the rat. Indeed, those differences reported or discussed at recent meetings involve either small numbers of substitutions (as seen between *mtp2* alleles of identical *cim* type) or truncation at the carboxy terminus²⁰, the region which is thought not to contribute to the membrane channel and which bears sequence similarity to the ATP binding site of other transporters.

Laboratory rats derive from the brown or Norway rat (*Rattus norvegicus*). Despite the name, this species originated in the Orient where it adapted to living with humans. By the eighteenth century it had spread to Europe and North America, generally replacing the earlier migrant, *Rattus rattus*²¹. Whereas humans and mice have three classical class I loci for presenting antigens to T cells, rats have just the one, *RT1.A*, and its juxtaposition to the class II region, like H-2K in mice, is distinct from that seen for the human class I genes¹⁶. Also different is the polymorphism of *RT1.A* — it is estimated that most polymorphic class I genes of humans and mice have over 100 alleles, yet analysis of 110 inbred rat strains has yielded just 14 serological *RT1.A* alleles and sampling of wild rats in Europe and North America did not extend that number²².

Single genes encode the two subunits of the putative transporter in rats, humans and mice. In the rat, co-evolution of transporter and class I alleles presents the least complicated proposition as each transporter has to supply only a single class I locus. In humans and mice, however, many mouths need feeding. A potential response to this more complex situation is co-evolution of multiple class I loci through the action of intergenic conversion, and for mice there is evidence for this mechanism²³. On the other hand, in humans the class I loci are relatively divergent and intergenic conversion seems to have been less of an evolutionary force than in mice²⁴. In this case the transporters may have become coupled to certain loci and uncoupled from others, possibly explaining the poor efficiency in assembly and cell-surface expression of HLA-C molecules compared to HLA-A and B²⁵. Another possibility is that functionally linked polymorphisms between transporter and class I alleles have not been elaborated in humans, as perhaps suggested by the evidence for recombination between class I and transporter genes.

That polymorphic genes in the class II region control antigen presentation by

class I molecules creates an ironic twist for the study of HLA and disease. Disease susceptibilities and histocompatibility determinants were first mapped to the class I loci. Subsequently, however, many associations proved stronger with alleles of the class II loci, class I then becoming more a genetic flag. As a consequence, research on HLA-associated disease and transplant rejection became infused with conviction that class II antigen presentation is at the root of most evil. Faustman *et al.*, in a much criticized paper²⁶ on diabetes, recently drew attention to an alternative mechanism also alluded to by Powis *et al.* This scheme raises class II as flag for the linked transporter polymorphisms which cause disease via class I-mediated presentation of antigens to CD8⁺ T cells. It is well known that CD8⁺ T cells are specialists in killing, while HLA-associated diseases often cause destruction of a particular target tissue. Could they possibly be related? □

Peter Parham is in the Departments of Cell Biology, and Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA.

VOLCANOES

You can pile it only so high

Paul T. Delaney

ALTHOUGH Mount Etna is currently in the headlines because of the threat posed by lava flows approaching villages on its flanks, a different and less expected hazard is described by Borgia *et al.* on page 231 of this issue¹. This great volcanic mountain is built on top of soft clay. Like the house built upon sand, Etna, without a solid foundation to support its enormous weight, is collapsing near its summit and spreading at its flanks. The southeastern portion of Etna's edifice, which is the site of its largest recurrent earthquakes, has been creeping outward at a rate of a centimetre or so each year; it may be dangerously unstable.

Until recently, most geologists regarded the fundamental hazards presented by volcanic activity to be those produced by the magmas that feed eruptions: lava and pyroclastic flows, ash falls and rain-induced volcanic mudslides (lahars). Additionally, craters and calderas, prominent on many volcanoes, are formed by particularly explosive activity and by foundering and collapse along ring faults as magma is withdrawn from its reservoir below. Caldera-like morphologies of some volcanic summits are now known to be due to landslides

caused by the weak and poorly lithified character of many volcanic rocks. The lesson was learned by all in 1980 when unloading of gas-charged magma by landslide removal of almost 3 km³ from the upper slopes of Mount St Helens triggered its spectacular eruption.

This 'sector collapse' produced a large scallop in the upper kilometre of the stratovolcano and left a trail of debris along 25 km of the valley below. As Seibert described in these columns recently², dozens of debris-avalanche deposits have now been recognized on and around volcanoes worldwide with run-outs of up to 125 km. With that has come the realization that volcanic edifices can be as hazardous as eruption of the magma contained within.

Previous workers^{3,4} recognized the landslide-style morphology of the upper Valle de Bove, a 7-km-long scallop-shaped valley southeast of Etna's summit, and even described features that are easily interpreted as the 'headwall' of a large landslide that forms the top of the valley. But Borgia and his colleagues draw attention to clay-rich deposits and old lavas flanking Etna from the south and east. The clay-rich rocks extend beneath Etna, forming its foundation,

- Deverson, E. V. *et al.* *Nature* **348**, 738–741 (1990).
- Trowsdale, J. *et al.* *Nature* **348**, 741–744 (1990).
- Spies, T. *et al.* *Nature* **354**, 744–747 (1990).
- Monaco, J. J., Cho, S. & Attaya, M. *Science* **250**, 1723–1726 (1990).
- Powis, S. J. *et al.* *Nature* **354**, 528–531 (1991).
- Spies, T. & DeMars, R. *Nature* **351**, 323–324 (1991).
- Kelly, A. *et al.* *Nature* **355**, 641–644 (1992).
- Spies, T. *et al.* *Nature* **355**, 644–646 (1992).
- Attaya, M. *et al.* *Nature* **355**, 647–649 (1992).
- Dobberstein, B. *Nature* **355**, 109–110 (1992).
- Parham, P. *Nature* **348**, 674–675 (1990).
- Parham, P. *Nature* **351**, 271–272 (1991).
- Robertson, M. *Nature* **353**, 300–301 (1991).
- Powis, S. J. *et al.* *Nature* **357**, 211–215 (1992).
- Livingstone, A. M., Butcher, G. W. & Howard, J. C. *Transplant Proc.* **15**, 1557–1559 (1983).
- Livingstone, A. M. *et al.* *J. exp. Med.* **170**, 777–795 (1989).
- Livingstone, A. M. *et al.* *Immunogenetics* **34**, 157–163 (1991).
- Hughes, A. L. & Nei, M. *Nature* **335**, 167–170 (1988).
- Pazmany, L. *et al.* *J. exp. Med.* **175**, 361–369 (1992).
- Powis, S. H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **89**, 1463–1467 (1992).
- Robinson, R. in *Evolution of Domesticated Animals* (ed. Mason, I. A.) 284–290 (Longman, Harlow, 1984).
- Cramer, D. V. *et al.* *J. Immun.* **120**, 179–187 (1978).
- Pease, L. R. *et al.* *CRC Crit. Rev. Immun.* **11**, 1–32 (1991).
- Parham, P., Lawlor, D. A., Lomen, C. E. & Ennis, P. D. *J. Immun.* **142**, 3937–3950 (1989).
- Stam, N. J., Spits, H. & Ploegh, H. L. *J. Immun.* **137**, 2299–2306 (1986).
- Faustman, D. *et al.* *Science* **254**, 1756–1761 (1991).
- Bodmer, J. G. *et al.* *Tissue Antigens* **39**, 161–173 (1992).

and are deflected downward by its weight. Where exposed on Etna's flank, however, they have been subjected to uplift, folding and thrust faulting. Younger volcanic rocks on the lower south-eastern slopes of the volcano have been similarly folded and faulted. These structures almost complete the picture of a large landslide mass that slips away from Etna's summit and accumulates down-slope towards the coast 20 km away. Indeed, although the crucial bathymetric data are lacking, it appears that landslide and creep-related deposits extend another 10 km or so southeastward along the floor of the Ionian sea.

The presence in the folds and thrusts of rocks that form the foundation on which Etna has been built provides strong evidence that the entire southern and eastern edifice is involved. Etna cannot support its own weight and spreads by landsliding of rocks high on the edifice and squeezing of rocks near its base as it grows through successive eruptions.

This interpretation of Etna's structure arrives on the heels of startling discoveries of submarine landslide deposits that flank oceanic volcanoes, especially along the Hawaiian island chain^{5,6} and at La Reunion, in the Indian Ocean⁷. Like Etna, these volcanoes (which rise as much as 9 km above the sea bed; Etna is about 3 km high) are built over soft basement rocks, in this case siliceous oceanic sediments. The Hawaiian submarine deposits are reckoned to be the largest landslides on Earth. Some extend for almost 250 km along the sea floor of the Hawaiian Deep; all are traceable back to the steep slopes of the volcanoes from which they slid. The larger deposits are estimated to have volumes of about 5,000 km³; for a typical volcano, this implies that the equivalent of a block 20 km by 20 km or more in area and the full depth of the edifice has been removed. Indeed, the larger landslides are as voluminous as the largest flood-basalt and caldera-forming eruptions yet measured.

From morphological analysis of bathymetric, seismic reflection and side-scan sonar data, it seems that the slides vary from slowly moving slumps to rapidly moving debris avalanches that carried intact blocks the size of large buildings more than 200 km from their source. The largest landslides seem to recur every 100,000 years or so and they occur during all stages of growth of Hawaiian volcanoes. In 1984, Moore and Moore⁸ identified tsunami-like deposits almost 400 m above sea level on the Hawaiian island of Lanai which are strongly suggestive of an enormous prehistoric earthquake or mass-wasting event in the Hawaiian islands. We can now see that the deposit was probably caused by the most recent of the Hawaiian mega-

GIANT volcanic slumps are not confined to Earth. Olympus Mons, on Mars, is the largest known volcano in the Solar System. It is 27 km high and over 600 km wide; the summit caldera (central depression) is 3 km deep and 80 km across. Olympus Mons is surrounded by a prominent, outward-facing scarp that is as much as 8 km high. Lava flows drape this scarp in places. The origin of the perimeter scarp has long been



controversial. Borgia *et al.*¹¹ have argued, by analogy with well-studied terrestrial volcanoes, that the scarp may have formed by motion on a system of thrust faults that intersect the Martian surface from beneath Olympus Mons, so that its lower slopes have been pushed upward and outward as the upper slopes and summit have settled and extended. If so, it provides a spectacular example of volcanic edifice instability. Some confirmation of this interpretation was provided by Thomas *et al.*¹², who examined terrace-like features further up the slopes of Olympus Mons (not visible on the accompanying image) which they also interpreted as thrust faults. Not content with an interpretation based upon examination of images, these workers calculated the stresses caused by a rock mass of the form of Olympus Mons. They showed that there are shear stresses along the middle slopes of the volcano much larger than those required to drive thrust motion. Further, they showed that these stresses would be modulated by the presence of a magma chamber within the volcanic edifice and that shear stresses appropriate to drive thrust motions could extend to the base of the volcano. Similar features have been identified on other Martian volcanoes. (Image was produced by merging the Mars medium-resolution Digital Image Model with low-resolution colour observations at the US Geological Survey in Flagstaff, Arizona. Image processing by Annie Allison and Alfred McEwen.) **P.T.D.**

slides, the Alike slide, which slipped off the west flank of Mauna Loa Volcano 105,000 years ago. Earlier this year, Young and Bryant⁹ proposed that erosive features found along the sea shore of southeast Australia, 7,000 km away, were produced by the same giant wave.

Fortunately, modern examples of volcanic edifice instability similar to those in Hawaii or at Etna are characterized for the most part, by creeping motion. These movements were identified in the 1960s, when high-precision ranging instruments were first used to measure surface displacements on volcanoes. Initially these measurements were intended to monitor accumulation of magma during swelling of volcanoes before eruption. Many shield volcanoes, however, contain rift zones formed by repeated injection of magma in long extensional cracks (dikes) and these quickly gained attention. At both Kilauea Volcano in Hawaii and Piton de la Fournaise in La Reunion, as well as Etna, it has been argued that the pressure of magma in dikes pushes the edifice outward, causing the over-steepening that results in slumping and landsliding. The depth to

which the Hawaiian and Reunion landslides extend is not well determined, however, because the volcanoes are partially — in fact, largely — submerged. A strong indication of that depth is provided by earthquakes, the largest of which originate deep within the flanks of these volcanoes. If comparison with Etna is valid then the answer is that landslide-like structures extend right to the base of these edifices as well.

The force balance among the weight

1. Borgia, A., Ferrari, L. & Pasquare, G. *Nature* **357**, 231–235 (1992).
2. Seibert, L. *Nature* **356**, 658–659 (1992).
3. Guest, J. E., Chester, D. K. & Duncan, A. M. *J. Volc. geotherm. Res.* **21**, 1–23 (1984).
4. McGuire, W. J. *J. Volc. geotherm. Res.* **26**, 377–386 (1985).
5. Lipman, P. W. *et al. J. geophys. Res.* **93**, 4279–4299 (1988).
6. Moore, J. G. *et al. J. geophys. Res.* **94**, 17,465–17,484 (1989).
7. Lenat, J.-F., Vincent, P., & Bachelery, P. *J. Volc. geotherm. Res.* **36**, 1–36 (1989).
8. Moore, J. G. & Moore, G. W. *Science* **226**, 1312–1315 (1984).
9. Young, R. W. & Bryant, E. A. *Geology* **20**, 199–202 (1992).
10. Dieterich, J. H. *J. geophys. Res.* **93**, 4258–4270 (1988).
11. Borgia, A. *et al. J. geophys. Res.* **95**, 14,357–14,382 (1990).
12. Thomas, P. J., Squyres, S. W. & Carr, M. H. *J. geophys. Res.* **95**, 14,345–14,355 (1990).

of a volcanic edifice, magma pressure, steepness of the upper slopes and buoyancy of groundwater, on the one hand, and resistance to slip, on the other, is not easy to judge. The order-of-magnitude analysis of Dieterich¹⁰, however, suggests that edifice height is the fundamental variable in determining the resistance to sliding because it reflects both an increase in the surface slope and in the overall driving pressure of the magma. In this regard, the flank earthquake of 1975 at Kilauea, which registered 7.2 on the Richter scale and is one of the largest to have struck a volcano this century, is particularly interesting. It produced about 8 m of down-slope displacement and occurred when Kilauea's edifice stood higher than at any time since the early 1920s.

The new interpretation of Etna is the latest in what amounts to an assault on

the traditional view of volcanic structure. Volcanoes, particularly those with steep slopes and those that extend below sea level, are likely to have internal features that reflect numerous surfaces of slip and large regions of internal instability. It is not only that debris avalanches, like that at Mount St Helens, pose a threat, but that movement of rocks near the base of volcanic edifices can generate earthquakes far more powerful than those caused by concomitant accumulation and migration of magma within the edifices. The oceanic setting of many of the suspect volcanoes means that the earthquakes generate tsunami-like waves which can damage coastal areas far from the source of disturbance. □

Paul T. Delaney is at the US Geological Survey, 2255 North Gemini Drive, Flagstaff, Arizona 86001, USA.

ENZYME COMPLEXES

A farewell to arms

David J. DeRosier

LIKE the Venus de Milo, the model of the dihydrolipoyl transacetylase (E2p) recently produced by Mattevi *et al.*¹ is an armless statue of pleasing symmetry. Mattevi *et al.* removed the arms (the amino-terminal 381 amino acids of the 638-residue E2p subunit) which each contain three lipoyl attachment domains and an E3 binding domain. The armless body is a cluster of 24 identical domains arranged at the vertices of a truncated cube. With the arms intact, the body can be clothed with copies of two different enzymes (E1p and E3). The complete structure, the pyruvate dehydrogenase enzyme complex, produces acetyl CoA from pyruvic acid. E2p catalyses one step of the reaction: it joins the decarboxylated pyruvic acid moiety to CoA. Lester Reed, who discovered the complexes, proposed that the lipoyl-containing arms swing, carrying substrates from one catalytic site to the next, so that each site can perform its reaction².

Why crystallize and analyse an armless body? Segments of flexible polypeptide tether the arm's four domains to the E2p body. In the electron microscope, a fuzzy exterior arising from the flexibly tethered domains surrounds the cube-like form of E2p^{3,4}. NMR spectra reveal the chaotic motion of the flexible linkers in solution⁵. With its flexible and therefore disordered arms, E2p resists crystallization, but with the arms removed it crys-

tallizes readily. From their work on the crystals, Mattevi *et al.*¹ have been able to study the structure, catalytic mechanisms and subunit interactions.

From their similarity in amino-acid sequence, the armless E2p should resemble chloramphenicol acetyl transferase (CAT), a trimeric enzyme that *ortho*-acetylates chloramphenicol using acetyl-CoA⁶. CAT has structural features of particular interest: a tunnel at the cataly-

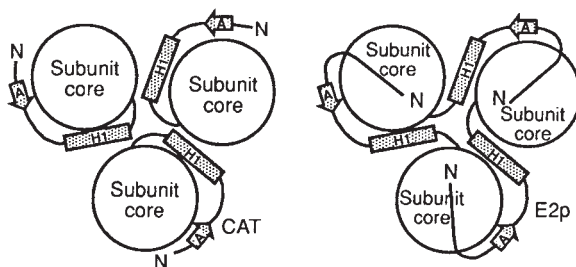


Diagram of the CAT (left) and E2p trimers (from ref. 1). The stump of the armless E2p and the amino-terminal 25 residues of CAT possess a sequence that forms a β -sheet (A) and a sequence that forms an α -helix (H1). In CAT, the amino-terminal sequence folds back onto the carboxy-terminal remainder of its own subunit, whereas in E2p this sequence loops over to interact with a neighbouring subunit thus lacing together the subunits of the trimer.

tic site with a critical serine and an unusual histidine. The histidine twists around so that the NH of its imidazole ring can hydrogen bond to its own carbonyl oxygen. These residues figure importantly in the catalytic mechanism. The tunnel, serine and histidine are also present in the E2p structure, suggesting that the two enzymes have a common

catalytic mechanism.

Although their catalytic mechanisms seem to be similar, the trimer intersubunit interactions differ (see figure). In CAT, the amino-terminal segment folds back onto the carboxy-terminal remainder of its own subunit. In the armless E2p, however, the homologous amino-terminal stump of one subunit loops over to complete a β -sheet in a neighbouring subunit. These interactions lace up subunits into trimers that occupy the vertices of the E2p cube. Such interactions are reminiscent of the interactions in simian virus 40, in which the carboxy-terminal tentacle of the main capsid protein, VP1, completes a β -sheet domain in a neighbouring subunit^{7,8}. Such interlocking intimately associates the folding process with subunit assembly and provides for stable interactions between subunits. Perhaps the E2p trimer requires more stability than the CAT trimer, because attached to the E2p are the swinging arms with their lipoyl moieties and E3 binding sites. The catalytic site shared by subunits would be disrupted if the trimer came undone.

The armless symmetric E2p complex fits in with the ideas proposed by Watson and Crick⁹ for viruses: structures made of identical subunits will exhibit a well-defined symmetry and stoichiometry. The E2p core has a well-defined symmetry and one can imagine that the remaining two enzymes would bind to it with the same symmetry and stoichiometry. Such a complex would contain 24 identical E1p-E2p-E3 clusters, each of which would have the same stereospecific interactions and catalytic sites.

Many multisubunit complexes use these principles. The pyruvate dehydrogenase enzyme complex (PDC) is not one of them.

The symmetry of PDC may be lower than that of its E2p core because its subunit stoichiometry is not 1:1:1, but 1:1:0.5. In other oxoacid dehydrogenase complexes, the stoichiometries deviate even more from 1:1:1. The complexes, however, function efficiently without high symmetry or equimolar stoichiometry because, with their swinging arms, they pass acetyl groups around the complex by a rapid intracomplex transacylation^{10,11}. In this way, an E1p can provide acetyl groups to E2p subunits in distant

parts of the complex. With the freedom provided by the arms, the complex achieves specificity without the need for rigidity or symmetry. Thus E1p and E3 must bind specifically to their sites on the E2p core but they need not bind in a symmetric fashion. Because there is more than one way to arrange the six E3 dimers on the E2p core, PDC may be

heterogeneous in its quaternary structure. Indeed, for the oxoglutarate dehydrogenase complex, Wagenknecht *et al.*¹² suggested that there are over 10^5 ways to bind E1 α and E3 to E2 α .

Given a heterogeneous population of complexes with arms and attached E3 subunits that swing and move about, how can one hope to solve the 'structure'? Not one or even two methods can do the job completely. But three can. By classical X-ray crystallography, Mattevi *et al.* solved the structure of crystals of E3 (ref. 13) and crystals of the armless E2p core¹. X-ray crystallography can produce images of the isolated components, but it cannot see parts that flap and swing, so that even if the arms were still attached to E2p, the electron density maps would have been as armless as the statue of Venus. Paradoxically, NMR can see the flapping arms on the E2p core but it cannot see the core itself. NMR continues to provide structural data on the swinging arms, and the lipoyl binding domain³, and the E3 binding domain¹⁴. But even X-ray crystallography and NMR together cannot see the whole complex.

That more difficult task falls to the electron microscope, which requires neither crystals, homogeneity nor high mobility. By fashioning a model of the complex from the known atomic structures and available electron micrographs, Mattevi *et al.* bring us closer to our goal of visualizing the whole assembly. I look at their orderly model, but in my mind I see a disordered E2p with the arms and clothes of E1p and E3 swinging and flapping about as they shuttle substrates from one site to the other. "Delight in disorder" by Robert Herrick (1591–1674) says it best:

A sweet disorder in the dress

*Do more bewitch me, than when art
Is too precise in every part.*

David J. DeRosier is in the Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254, USA.

Stars arriving two by two

Cathie Clarke

MOST stars reside in binary systems, and, according to recent research, many are observed to be paired up at the earliest stages of stellar evolution. In fact, new results reported at a colloquium last month* suggest that binaries may be even more common amongst pre-main sequence stars than amongst their older, main sequence counterparts. This implies, at the very least, that the formation of stellar pairs is inextricably linked with the initial processes of star formation, and that any successful theory of star formation must predict this preference of stars for the paired state. Indeed, taken at face value, this apparent excess of pre-main sequence binaries implies that even single stars on the main sequence (such as the Sun) might have originated in binaries. If so, what caused some pairs to part company?

A basic lesson of the colloquium was the high scientific value of statistical observational projects — those that accumulate the orbital parameters of large numbers of binaries. Counting binaries is a much under-rated activity in the astronomical community at large, in great part because it offers no quick return on invested effort, requiring instead repeated, precise observations over a long period. Thus catalogues of visual binaries, those in which the individual components can be separately resolved, such as that of the US Naval Observatory, draw on lists extending back to the eighteenth century. Spectroscopic searches (such as that using the high-precision CORAVEL spectrometers at Haut-Provence and La Silla Observatories) require repeated measurements over periods of at least ten years to track down the movements of components which cannot be seen individually.

These traditional techniques are now being supplemented by interferometric methods, such as the speckle survey under way at the Center for High Angular Resolution Astronomy at Georgia State University, which has detected thousands of binaries with intermediate separations. And a great hope for the future lies in long-baseline optical interferometry (using arrays like the Sydney University Stellar Interferometer), which offers the prospect of resolving even the closest binaries. Taken together, the databases acquired through these diverse techniques can now begin to answer a number of questions concerning the properties of binary stars, and thus pro-

vide crucial clues for those trying to devise theories for binary star formation.

The picture that is starting to emerge from this wealth of data is that binaries are very common among main-sequence stars, current estimates suggesting that the fraction of binaries among solar type stars is at least 65 per cent (M. Mayor, Geneva Observatory). These binaries span six orders of magnitude in separation, from systems that are almost in contact to those that are so wide that they are destined to be dissolved eventually by encounters with field stars. The orbits are generally highly eccentric, implying that binary companions must form in a fundamentally different way to planets, whose orbits are pretty much circular. One notable result is an apparent dearth of the highest eccentricities among systems with closer separations, implying that the binary formation process cannot be entirely independent of scale. This effect extends to far larger separations than can be accounted for by orbital circularization, which arises from tidal effects between the two stars if they get sufficiently close to each other. This tidal circularization of the very closest binaries is reasonably well understood, but the lack of high-eccentricity orbits in somewhat wider systems remains a puzzle.

Observational controversy still surrounds the mass spectrum with which stars pick their partners: the most recent re-evaluation of spectroscopic data has lent support to an old idea that whereas in wider pairs, partners are picked at random, the close pairs show a preference for more nearly equal mass components (T. Mazeh, Tel Aviv University). If correct, this distinction would indicate that separate formation mechanisms might operate for the close and wide pairs — again, an essential input to theories of binary formation.

Running alongside the establishment of this picture of main-sequence binary properties has been the exciting discovery of large numbers of pre-main sequence binaries, that is binaries in their stellar youth. This is a rapidly expanding area, no spectroscopic binaries having been discovered among pre-main sequence stars before 1983. Now more than twenty of these systems are known, and they number amongst them some of the youngest stars that are optically visible. Such stars are probably less than a million years old, so that binaries can evidently form very early in the stellar evolutionary process (R. Mathieu, University of Wisconsin). Imaging surveys of emission-line stars in nearby dark clouds

1. Mattevi, A. *et al.* *Science* **255**, 1544–1550 (1992).
2. Reed, L. J. *Acc. chem. Res.* **7**, 40–46 (1974).
3. Biele, D. M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **76**, 4385–4389 (1979).
4. Wagenknecht, T. *et al.* *J. biol. Chem.* **265**, 22402–22408 (1990).
5. Perham, R. N. *Biochemistry* **30**, 8501–8512 (1991).
6. Guest, J. R. *FEMS Microbiol. Lett.* **44**, 417–422 (1987).
7. Liddington, R. C. *et al.* *Nature* **354**, 278–284 (1991).
8. Caspar, D. L. D. *Curr. Biol.* **2**, 169–171 (1992).
9. Crick, F. H. C. & Watson, J. D. *Nature* **177**, 473–475 (1956).
10. Bates, D. L. *et al.* *Nature* **268**, 313–316 (1977).
11. Collins, J. H. & Reed, L. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4223–4227 (1977).
12. Wagenknecht, T. *et al.* *J. molec. Biol.* **165**, 523–541 (1983).
13. Mattevi, A. *et al.* *J. molec. Biol.* **220**, 975–994 (1991).
14. Robien, M. A. *et al.* *Biochemistry* **31**, 3463–3471 (1992).

*Complementary Approaches to Double and Multiple Star Research, IAU Colloquium 135, Pine Mountain, Georgia, 5–10 April 1992.

are turning up a number of wide binaries in regions of star formation (H. Zinnecker, University of Wurzburg). It is, however, the use of interferometric methods, sensitive to binaries of intermediate separations, that has so vastly expanded the roster of pre-main sequence binaries, particularly through near infrared-speckle and lunar-occultation techniques. (The use of the near infrared passband allows the detection of companions that are not yet optically visible, being still shrouded in their natal gas and dust). It is among these systems, with separations in the range ten to several hundred astronomical units (the Earth's orbit has a radius of one astronomical unit), that there would appear to be an excess of binaries compared with main-sequence counts, according to reports from three independent groups (A. Ghez, California Institute of Technology; C. Leinert, University of Heidelberg; M. Simon, State University of New York at Stony Brook). With the present sample sizes this result is significant only at the two sigma level. But if it persists in larger samples it poses a number of questions: either a large number of systems are being missed in main-

sequence surveys or else many binaries in this separation range are somehow destroyed during their main-sequence lifetimes.

Whatever the outcome of this particular controversy, the discovery of so many pre-main-sequence binaries provides a welcome dataset against which theorists can test their pictures of binary star formation. For instead of having only to produce a correctly finished product, such models have now also to be checked against the appearance of binaries that are still in the late stages of formation — in which, for example, the remnant gas around the young binary can provide a clue as to the initial formation process. Thus numerical models, such as those of A. Boss (Carnegie Institute: see *Nature* **351**, 274; 298–300; 1991) and I. Bonnell (University of Montreal) have now to pass more stringent observational tests than ever before. And this can only hasten the arrival of a consensus about binary formation — and thus, implicitly, about all star formation. □

Cathie Clarke is in the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.

EVOLUTIONARY CHEMISTRY

Life in a test tube

Laurence D. Hurst and Richard Dawkins

How many lives are there? How many lives could there be? Not lives, lifes. We have experience of only one life, the one on this planet based on DNA as replicator and protein as executor. Is DNA the only molecule with the necessary qualifications? Are there others but vanishingly rare, such that the origin of a life in the Universe is a prodigiously lucky event? Or is there a profusion of alternative biochemistries waiting to be discovered? Work by a group led by Julius Rebek, described in a forthcoming paper¹ and in this week's *Science*², suggests that not only can relatively small molecules act as replicators but that it is quite realistic to consider whole other-worlds of chemical replicators.

Rebek's first test-tube replicator had two basic components³ — an imide ester (which we will call molecule A) and an adenine-containing amine (molecule B). In a solvent the two molecules pair to form a third molecule (molecule C). It is this third molecule that acts as a replicator. Molecule C not only acts as a template on which A and B are guided by hydrogen bonds to line up, but also catalyses the covalent union between A and B. Two identical C molecules then become available to carry on the autocatalysis. Just as a laboratory

population of bacteria exponentially grows until it runs out of resources, so growth of the population of C molecules describes the classic sigmoidal form.

As it stands, however, selection (and hence evolution) in this system is impossible as there are no variants. By the incorporation of a variety of B-type molecules which differ only in single side groups, the system can be modified to take account of this⁴. The resulting different C-type molecules have both differing replicating (autocatalytic) capabilities, as well as differing capabilities to catalyse the formation of non-like C-type molecules. A population of two rival variants of C molecules now exists in a state of competition. This variation had to be artificially created, but one of the new B types made the C molecules sensitive to ultraviolet radiation. Exposure to ultraviolet cleaves off part of the B subunit of C, producing a new C-type molecule. This new molecule not only replicates, it actually outcompetes the unmutated molecule and rapidly takes over the system's resources⁴. This then is a replicating system in which new variants are spontaneously generated, albeit in a rather limited fashion.

In their paper in press with *Journal of the American Chemical Society*¹, Rebek

et al. describe a wholly different chemical replicator. This too has two subunits, D and E, which pair to form a new molecule, F. Like C, F can catalyse the synthesis of more F from D and E. Mutation in this system has yet to be described.

The next move was to attempt to make hybrids² and to investigate whether A with D or B with E would make a replicator. The team comment that "at first glance both recombinants might be expected to replicate. They both bear self-complementary recognition surfaces, and can gather their respective reaction components in termolecular complexes". Duly, one of the hybrids did replicate, and in no insignificant terms. It was in fact the most efficient replicator found to date. Conversely, the other was sterile. The difference between the two was simply a question of the orientations of the respective recognition surfaces. The successful hybrid had the recognition surfaces in parallel whereas the other hybrid was either C shaped or S shaped. In neither the C nor S form could the stable intermediate structure be formed.

What does this elegant system tell us about evolution and the origin of life? At the very least it shows that DNA and RNA are not the only possible replicators. However, Rebek's work might have another importance. Short of travelling the Galaxy, it is hard for us to know which features of life are particular, which general. We can surmise that some form of darwinian selection is universal for all lifes⁵ but beyond that we cannot say what chemical properties would be necessary. Rebek's work allows us to approach the question of which chemical and structural aspects of replicators are necessary, and which are details. We can for instance start to ask whether some types of structures are inherently more or less likely to become replicators, and whether there are general properties of all of the chemical structures which can replicate. The new studies show that replicators need not be particularly large molecules and that simple replicators could have the potential to mutate. Life could have relatively modest beginnings. Similarly, the fact that the C- and S-form hybrid molecules were sterile is significant in that it goes some way to defining necessary structural characteristics which replicators must have. Furthermore, Rebek's chemicals and DNA share a number of properties: they are organic, use hydrogen bonding as structural information and covalent bonds to maintain structural integrity. But until we have a larger compendium of replicators we will not know if these are necessary characteristics or simply characteristics which happen to be shared by these two systems.

ELECTRONICS

Single atoms as transistors

Sean Washburn

Although Rebek's system is unique in employing synthetic chemicals, analogous systems of replicating molecules have been set up using biological chemicals⁶. Whereas these biological replicators can be seen as attempts to understand what actually occurred at the origin of life, Rebek's system, like Thomas Ray's replicatory computer world^{7,8}, can be seen as a different, but parallel, replicating kingdom. Ray's work might best be thought of as being complementary to that of Rebek. Where Rebek's programme attempts to investigate the conditions necessary for the initial evolution of a life, Ray's replicating 'information sets' are defined with the ability to replicate and hence permit study of the consequences of mutation and selection in a population of simple replicators. The revealing aspect of Ray's work is that a diverse population can evolve from such an initial population of simple replicators. For instance, the tendency for parasites (replicators which use another's replicatory machinery) to invade and persist means that parasitism might be an unavoidable aspect of all lives.

There might however be a fruitful connection between the two approaches. Just as we can ask whether particular types of 'information sets' differ in their evolutionary potential, so we can now ask whether different systems of chemical replicators differ in their capacity to evolve, and if so why. Where Rebek's system (and probably protein replicators^{9,10} as well) falls short is that these replicators, unlike DNA, do not encode much information. This must surely limit their evolutionary potential. But it is probable that in early evolution DNA (or whatever the early replicator was) didn't encode much information either. Could strings of Rebek-type chemicals evolve to encode more than their own replication? If so, would we want to think of this as a life in a test tube? □

Laurence D. Hurst and Richard Dawkins are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

As the drive to gigabit computer memory chips accelerates, requiring smaller and smaller transistors, one might ask what the limits are to making transistors smaller. Recent experiments indicate that the limit might be reached only at the level of single atoms. Three independent groups of researchers now show¹⁻³ that electrical current can be controlled by the configuration of an individual atom's quantum-mechanical state and

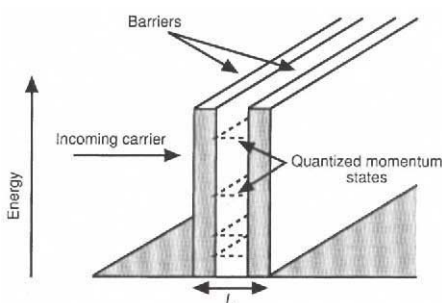


FIG. 1 Resonance tunnelling: a pair of barriers separate off a slice of semiconductor, a quantum well, inside which only discrete momentum states (broken lines) are possible. Charge carriers impinging from the left will be reflected unless their momentum is aligned with one of the allowed states in the well. Carriers in the well are free to move parallel to the plane.

that the state itself can be switched on and off.

The transistor of Dellow *et al.*¹, from the Universities of Nottingham and Glasgow, is based on the process of resonant tunnelling, which has been used for many years⁴, but never on so small a scale. Current through such a device is controlled by the existence of discrete quantum levels in a thin confined layer through which it must pass, whose energies can be adjusted externally (Fig. 1). Charge carriers entering from one side of the device cannot be transported across the barriers unless one of these levels in the quantum is tuned into resonance with the carriers' energy. To attain the quantum-well states, it is not necessary to restrict the charge-carriers' motion in all three directions. It is sufficient for the barriers to confine the motion along only the current direction on a scale of length which is short compared with the carriers' characteristic lengths, such as the wavelengths of their wavefunctions (a few nanometres in the GaAs semiconductor used by Dellow *et al.*¹).

So earlier work on resonant tunnelling concentrated on planar systems with motion free in the plane but confined across it⁵, or on tunnelling through localized

carrier states in disordered insulators⁶. If the motion is confined along other directions, the quantum effects can be more dramatic⁷. This drama occurs when the size w of the tunnelling region (Fig. 2a) is also comparable with the important characteristic length. Dellow *et al.* have used a constricting ring to squeeze the tunnelling area down to less than $0.1 \mu\text{m}^2$, inside which only a few donor atoms are active. The conduction region is schematically drawn in Fig. 2a, which shows that the carriers are confined to flow from the top through a small port of dimension w containing the two tunnel barriers to the bottom. Depending on the voltage settings on the transistor, most of the carriers are funnelled through only a single donor. By changing the voltage settings, the authors can cause the carriers to flow through one or another of the donors or study the voltage dependence of the donors' states. As the donors move through the resonance, peaks appear in the transistor's current-voltage curve. The peaks mark the energy spectrum of the donors.

Gregory, of Bellcore, has also studied tunnelling of electrons through a single atom between two wires that nearly touch². His ingenious technique is to bring the two thin tungsten wires very close to each other magnetically, by passing a current through one in an applied field, and to cement them in place with the van der Waals forces between helium atoms adsorbed onto their surfaces. In such a configuration, the electrons tunnel from wire to wire preferentially through the nearest two points (as the tunnelling probability is exponentially smaller elsewhere). Again, a single impurity atom can be the funnel point, just as a particular donor was in the device of Dellow *et al.* (Fig. 2b).

Gregory used this technique to study tunnelling through magnetic impurities. Sometimes there is a magnetic inclusion in the asperity where tunnelling occurs. This leaves a specific signature in the current-voltage curve, and, because the interaction with the carriers is magnetic, the signature can be altered by a magnetic field. The signature is a 'zero-bias-anomaly' similar to those studied for several decades^{8,9}. Tunnelling experi-

Correction

It has been pointed out that increased error estimates, mentioned in a Résumé item (*Nature* **356**, 288; 1992) on the latest γ -ray burst statistics from the Compton Gamma Ray Observatory, are attributable to the inclusion of an estimate of the positional uncertainty in burst locations, and do not indicate deteriorating results.

1. Park, T. K., Feng, Q. & Rebek, J., Jr *J. Am. chem. Soc.* (in the press).
2. Feng, Q., Park, T. K. & Rebek, J., Jr *Science* **254**, 1179-1180 (1992).
3. Nowick, J., Feng, Q., Tijivikua, T., Ballester, P. & Rebek, J., Jr *J. Am. chem. Soc.* **113**, 8831-8839 (1991).
4. Hong, J.-I., Feng, Q., Rotello, V. & Rebek, J., Jr *Science* **255**, 848-850 (1992).
5. Dawkins, R. *The Selfish Gene* (Oxford University Press, 1989).
6. Zieliński, W. S. & Orgel, L. E. *Nature* **327**, 346-347 (1987).
7. Ray, T. S. in *Artificial Life II*, Santa Fe Inst. Studies in the Sciences of Complexity, Vol XI (eds Langton, C. G., Taylor, C., Farmer, J. D. & Rasmussen, S.) 371-408 (Addison-Wesley, Redwood City, California, 1991).
8. Maynard Smith, J. *Nature* **355**, 772-773 (1992).
9. Hurst, L. D. & Haig, D. *Nature* **351**, 21 (1991).
10. Cheng, M. Y., Hartl, F.-U. & Horwich, A. L. *Nature* **348**, 455-458 (1990).

ments with insulating barriers have been turning up these anomalies of raised conductance at low bias, which were difficult to explain (hence the vague, generic name). One explanation proposed is a magnetic interaction (the Kondo effect¹⁰) that involves the screening of a magnetic impurity by electron spins. In Gregory's experiment, the screening of a

single atom governing the conductance is either a donor fortuitously situated near the tunnel barrier¹ or a magnetic asperity on the surface of a tungsten wire². Although the experiments were carefully planned, the authors had to rely on luck for the placement of the impurities, which does not bode well for those who would like to make single-atom transistors.

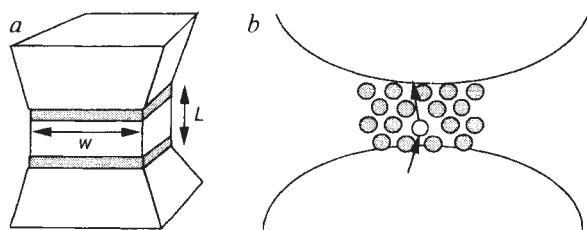


FIG. 2 The conduction path in a resonant tunnelling device can be restricted to a small port by (a) circuit lithography¹ or (b) tunnelling sensitivity² as in tunnelling microscopes. Impurities in the tunnelling regions almost wholly control the conductance if the port is small enough.

single magnetic atom apparently dominates the conductance, which can be made to change about tenfold. By shifting the wires, he can see discrete switching events that indicate sudden changes between different magnetic centres. (A change through many would be smooth.) The results of this experiment speak to a range of deep questions in physics of disordered materials, where magnetic interactions are at the heart of such problems as the transition from conductor to insulator¹¹.

How might such devices be used in electronic circuits? Resonant tunnelling is a straightforward extension of the Esaki (or tunnel) diode. The peaks in the current-voltage curve provide regions of negative conductivity — where current is increased by decreasing the bias voltage. In the Esaki diode, this region is used in making high-precision oscillator circuits, and the single-atom resonant-tunnelling devices could be used to the same effect. Alternatively, the confined region could be biased independently with a third electrode, to create something more like a conventional transistor, with the conductance between the in and out connections controlled by the third. With Gregory's device, it is hard to see how a third connection could be made, but the applied magnetic field, tweaking the zero-bias anomaly, would achieve the same purpose. It is unlikely that the crossed-wire system would be used as it stands, but it is a powerful tool (because of its mechanical stability) for studying tunnelling processes. Resonant tunnelling, on the other hand, already has a demonstrated range of applications in transistors¹² and further applications in logic circuits are proposed¹³.

In each of these experiments, the

single atom governing the conductance is either a donor fortuitously situated near the tunnel barrier¹ or a magnetic asperity on the surface of a tungsten wire². Although the experiments were carefully planned, the authors had to rely on luck for the placement of the impurities, which does not bode well for those who would like to make single-atom transistors. But several experiments have shown how one can use a scanning microscope probe to place atoms in specified positions on the surface of a semiconductor. Indeed, that was the basis of the atomic switch described by Eigler *et al.* last year³, in which a single xenon atom was toggled between 'on' and 'off' positions.

The question for the practical use of atom-scale electronics becomes one of quality control. If the atom-scale circuitry can be made to do the work of a transistor, then this is a small step towards its use in practical circuits. The problem is repeating the process reliably: this is already forbidding in contemporary electronics, where keeping production facilities clean has become one of the main manufacturing tasks. To pack memory cells one thousand times more densely and do it reliably will require herculean efforts to keep the manufacturing environment controlled.

This still leaves untouched the question of connecting the transistors together. Interconnects now take up a large fraction of the area on the largest modern chips. Some new scheme will be required for the atomic-scale transistors — the simple isolated transistor connected by wires to others will become a thing of the past, to be replaced by functionally integrated circuits dedicated to particular operations. □

Sean Washburn is in the Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3255, USA.

1. Dellow, M. W. *et al.* *Phys. Rev. Lett.* **68**, 1754–1757 (1992).
2. Gregory, S. *Phys. Rev. Lett.* **68**, 2070–2073 (1992).
3. Eigler, D. M., Lutz, C. P. & Rudge, W. E. *Nature* **352**, 600–603 (1991).
4. Chang, L. L., Esaki, L. & Tsu, R. *Appl. Phys. Lett.* **24**, 593–595 (1974).
5. Mendez, E. E. & von Klitzing, K. (eds) *Physics and Applications of Quantum Wells and Superlattices* (Plenum, New York, 1987).
6. Fowler, A. *et al.* *Phys. Rev. Lett.* **57**, 138–144 (1986).
7. Reed, M. A. *et al.* *Phys. Rev. Lett.* **60**, 535–538 (1988).
8. Wyatt, A. F. G. *Phys. Rev. Lett.* **13**, 401–403 (1964).
9. Logan, R. A. & Rowell, J. M. *Phys. Rev. Lett.* **13**, 404–406 (1964).
10. Applebaum, J. *et al.* *Phys. Rev.* **160**, 554 (1967).
11. Finkelstein, A. M. *JETP Lett.* **40**, 796–799 (1984).
12. Capasso, F. & Kiehl, R. A. *J. appl. Phys.* **58**, 1366–1369 (1991).
13. Sone, J. *Semicond. Sci. Tech.* **7B**, 210–214 (1992).

Raising the roof

A BALLOON is an elegant flying machine, but its huge envelope wastes a lot of space. Daedalus once proposed a Zeppelin filled, not with pure helium, but with 80% helium and 20% oxygen. This lifting gas supports respiration. The whole vast envelope could then double as a spacious cabin for crew and passengers — a welcome change from the cramped fuselage of conventional aircraft. With its low density and inertia, the gas would be wonderfully light and easy to breathe, a boon to asthmatics and bronchial sufferers. Sadly, it would also raise their voices to an unintelligible duck-speak.

The envelope of a hot-air balloon, full of propane combustion gases, would be less welcoming. It would need some sort of heat-exchanger, ejecting the fumes of combustion while transferring their heat to the entrapped air. Daedalus was glumly concluding that a balloon so burdened would never get airborne, when he realized the advantages of simply leaving it on the ground. An anchored hot-air balloon would make an ideal tent.

The conventional pressure dome, kept inflated by a continuously running blower, has to be well sealed around the rim, and must be entered through an airlock. By contrast, Daedalus's hemispherical 'Hot-house' will be kept inflated by the hot air inside it, balloon-fashion. Its base will be at atmospheric pressure; it will need no rim seal and can be entered through normal doors.

Like all balloons, Hot-houses will get better as they get bigger. Really large ones could afford quite a thick, insulating fabric construction. They would then stay inflated on relatively little heat — perhaps that released naturally by the inhabitants and their equipment. Cheap, colourful and effective, Hot-houses will be ideal for exhibition halls, sports-centres and even factories, especially in cold climates. In sunnier regions, transparent Hot-houses could be kept inflated by pure sunlight. They could then serve as huge, elegant agricultural greenhouses. But even a solar Hot-house will need some sort of thermostatted emergency heating system to keep it inflated on cold nights or to melt any snow-load that threatens to collapse it. And every Hot-house will need to be firmly anchored against taking off — but not too firmly. Daedalus's design has strong but easily melted ground couplings. If fire breaks out in a Hot-house, its restraining ties will melt. The whole envelope will simply rise into the air, taking the smoke and fumes of the blaze with it. The inhabitants can walk away unimpeded, while the fire services tackle whatever contents are burning on the ground.

David Jones

Original and artificial antibodies

SIR — Hybridomas immortalize B cells by cell fusion. The monoclonal antibodies derived by this method are made by what we refer to as an original pair of heavy (H) and light (L) chains. The situation is different in the case of antibodies derived by PCR cloning of the variable (V) genes followed by their expression in bacteria (reviewed in ref. 1) or on bacteriophages²⁻⁴. Current applications of these methods to bypass hybridomas involve separate amplification of the V_H and V_L segments, usually from messenger RNA and their random recombination during a single cloning step. Although antibody fragments with good affinity⁵ or good apparent affinity⁶ have been obtained in this way, it is unclear how often this approach leads to the recovery of the original pairs of H and L chains expressed by individual B cells⁷. It was predicted that libraries of 10⁸ H- and L-chain pairs or larger had to be constructed from B-cell mRNA in order to have a reasonable chance that rare original pairs could be rescued¹. This calculation has been called into question⁶, although the implication that it did not take into consideration the advantages of amplification of mRNA is incorrect (see Table 1 of ref. 1).

We have compared anti-2-phenyloxazalone (phOx) antibodies obtained by hybridomas with the already described⁵ single chain F_s (scF_v) made with mRNA from the same pool of spleen cells. The random combinatorial library (containing 2 × 10⁵ V_H and V_K gene pairs assembled and expressed on a filamentous phage) had been selected on a phOx affinity column and clones ranked by ELISA. One of these bound phOx

with a K_d = 10⁻⁸ mol⁻¹ (ref. 5). From the hybridomas derived from 5 × 10⁷ spleen cells, 43 positives were obtained. These were similarly ranked by ELISA and several clones among the strong and medium binders were characterized further. One of these clones (HD11.4) bound phOx with a K_d = 2.7 × 10⁻⁹ mol⁻¹ (see table). Thus, although both methods produce good binders, the best was isolated from a hybridoma. The gene combination of this antibody was not obtained from the random combinatorial library, and nor were other gene combinations typical of the anti-phOx response (see figure and table).

It has been suggested that random combinatorial libraries could be used to study the 'fossil record' of antibody responses⁸, but such a library, in our case, did not provide an accurate guide to the anti-phOx response. In part, this may be for technical reasons, for example selective PCR amplification (see table), or the single-chain format of the phage antibodies. No less important, however, is the promiscuity of H- and L-chain pairing in combinatorial libraries^{5,9}. Indeed, how do we identify a V_H-V_L original pair among all the promiscuous combinations which, at least in some cases, appear to dominate the combinatorial library? We suggest that for the analysis of B-cell memory, it is better to use V_H and V_L genes recovered from antigen selected cells (for example, ref. 10), and, for this purpose, amplification of sequences taken from DNA may be preferable to mRNA, because the latter is very biased towards plasma cells at the expense of the memory cells.

ORIGINAL ANTI-PHOX ANTIBODIES

Hybridoma	V _H segment	V _K segment	K _d (nM)
GG8.3*	V _H Ox1	V _K Ox1	66.9 ± 7.6
HS11.4*	V _H Ox2	V _K Ox1	2.7 ± 0.5
HG9.6	V _H 11	(V _K 45.1)	133.4 ± 20.8
IG10.1	V _H 11	(V _K 45.1)	48.6 ± 5.9

BALB/c mice were immunized with the hapten 2-phenyloxazalone and spleen cells from a pool of five animals were isolated three days after a secondary injection of the hapten. Aliquots of the same pool of cells were used for fusion with the mouse myeloma NSO and for mRNA extraction, PCR cloning of the V_H and V_K repertoires and phage expression⁵. By sequence analysis, 23 clones obtained from the combinatorial library appeared to be derived from eight individual V_H genes and seven V_K genes. Among them, the V_HOx1 and V_KOx1 genes, the gene combination that dominates the early response to phOx (ref. 12), were found only once and were not paired together⁵. By contrast, two of four of the hybridoma clones were somatic mutants of V_KOx1-V_HOx1 and V_KOx1-V_HOx2 pairings. The V_HOx2 gene is closely related to the V_HOx1 gene but appears less frequently than V_HOx1 in the phOx response. In combination with a mutated V_KOx1 gene, V_HOx2 gave rise to the highest affinity antibody HD11.4. Hybridomas HG9.6 and IG10.1 both used the V_H11 gene, probably in combination with V_K45.1 as commonly found in the secondary response to phOx. These genes were not rescued from the combinatorial library but this may be because the primers used do not amplify the V_K45.1 gene (our unpublished observation). The absence of V_H11 among the anti-phOx scF_v derived from the combinatorial library probably reflects the fact that V_H11 generates anti-phOx antibody in combination with V_K45.1 (ref.13).

High-affinity antibodies required for conventional radio-immunoassays constitute a very small fraction of the antibodies present in the serum of animals subjected to elaborate immunization protocols. Not surprisingly, monoclonal antibodies suitable for similar applications are also rare. Will random combinatorial libraries simplify the preparation of such antibodies? We suggest that the answer to this question depends on at least two factors. The first is the probability that 'non-original' pairs re-

a

VH0X1	CAG GTG CAG CTG AAG GAG TCA GGA CCT GGC CTG GTG GCG CCC TCA CAG AGC CTG TCC ATC ACT TGC ACT GTC TCT GGG TTT TCA TTA ACC AGC TAT GGT GTA CAC TGG GTT CGC CAG CCT	10	20	30	40
NQP35/GG8.3	--- --C--- --T--- --G--- --C---				
NQP35/HD11.4	--- --C--- --T--- --G--- --A--- --C--- --A--- --T--- --C--- --T---				

VH0X1	CCA GGA AAG GGT CTG GAG TGG CTG GGA GTA ATA TGG GCT GGT GGA AGC ACA AAT TAT AAT TCG GCT CTC ATG TCC AGA CTG AGC ATC AGC AAA GAC AAC TCC AAG AGC CAA GTT TTC TTA	50	60	70	80
NQP35/GG8.3	--- --G--- --A--- --T--- --A--- --C--- --C--- --G--- --T---				
NQP35/HD11.4	--- --G--- --A--- --G--- --G--- --A--- --T--- --A--- --C--- --C--- --G--- --T---				

VH0X1	AAA ATG AAC AGT CTG CAA ACT GAT GAC ACA GCC ATG TAC TAT TGT GCC AGA	82a 82b 82c	90
NQP35/GG8.3	--- *--- --G--- --C--- --A--- --T---		
NQP35/HD11.4	--- *--- --G--- --C--- --A--- --T---		

b

VK0X1	CAA ATT GTT CTC ACC CAG TCT CCA GCA ATC ATG TCT GCA TCT CCA GGG GAG AAG GTC ACC ATG ACC TGC AGT GCC AGC	10	20	27	34
NQP35/GG8.3	---	---	---	---	---
NQP35/HD11.4	---	---	---	---	---

VK0X1	TCA GGC ACC TCC CCC AAA AGA TGG ATT TAT GAC ACA TCC AAA CTG GCT TCT GGA GTC CCT GCT CGC TTC AGT GGC AGT	40	50	60	70
NQP35/GG8.3	---	---	---	---	---
NQP35/HD11.4	---	---	---	---	---

VK0X1	GCT GAA GAT GCT GCC ACT TAT TAC TGC CAG CAG TGG AGT AGT AAC CCA	80	90
NQP35/GG8.3	---	---	---
NQP35/HD11.4	---	---	---

Sequence to the V_H (a) and V_K (b) segments of GG8.3 and HD11.4 anti-phOx hybridomas and of the germline genes V_HOx1 and V_KOx1. A gap of two residues has been introduced, and marked with asterisks to indicate the segment where there are variations in the length of different light chains.

combine to produce antibodies of the same or higher affinity than the ones encoded by original V_H - V_L pairs. This is more likely to happen in experiments such as the one described in the table, dealing with early stages of hyperimmunization than when dealing with more elaborate hyperimmunization schedules designed to generate antibodies of the highest affinity. The second is the probability that a random combinatorial library may reconstruct the original V_H and V_L pair. Libraries containing 10^5 - 10^6 random pairs can clearly give rise to good antibodies but, at least in our case, (and as predicted in ref. 1) failed to reproduce known original pairs (table). This may not always be the case, especially as technology improves and the library size increases. However, as the complexity of the library increases, the problems associated with the recovery of individual pairs also increase. We consider that, short of more elaborate protocols, the recovery of such rare 'original' H and L chain pairs is critically dependent on very complex combinatorial libraries ($>10^8$) and well outside the capacity of plaque or colony-screening methods. It may be, however,

within reach of the phage technology^{2-4,11}, and this is highly relevant to human antibodies where hybridomas have not been easy to derive.

Ermano Gherardi*

César Milstein

MRC Laboratory of Molecular

Biology,

Hills Road,

Cambridge CB2 2QH, UK

* Present address: ICRF Cell Interactions Laboratory, Department of Medicine, University of Cambridge, Cambridge, CB2 2QQ UK.

1. Winter, G. & Milstein, C. *Nature* **349**, 293-299 (1991).
2. McCafferty, J. A. *et al. Nature* **348**, 552-554 (1990).
3. Kang, A. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 4363-4366 (1991).
4. Breitling, F. *et al. Gene* **104**, 147-153 (1991).
5. Jackson, T. *et al. Nature* **352**, 624-628 (1991).
6. Burton, D. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 10134-10137 (1991).
7. Caton, A. J. & Kaprowski, H. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6450-6454 (1990).
8. Lerner, R. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 9705-9706 (1991).
9. Kang, A. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 11120-11123 (1991).
10. Rada, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 5508-5512 (1991).
11. Smith, G. P. *Science* **228**, 1315-1317 (1985).
12. Kaartinen, M. *et al. Nature* **304**, 320-324 (1982).
13. Berek, C. & Milstein, C. *Immun. Rev.* **96**, 23-41 (1985).

Amino acids from coal gasification?

SIR — Relatively large amounts of two nonprotein amino acids, α -aminoisobutyric acid (AIB) and racemic isovaline (ISOVAL), have been identified in sediments proximate to the Cretaceous/Tertiary (K/T) boundary at Stevns Klint, Denmark¹. An alternative explanation for the occurrence of these amino acids which does not require extraterrestrial impacts is that they were formed as a result of coal gasification. This hypothesis is based on the isolation and identification by gas chromatography/mass spectrometry of large quantities of amino acid precursors (hydantoin) in gasification condensate water². The major hydantoins were 5,5-dimethylhydantoin and 5-ethyl-5-methylhydantoin², which are indeed those that generate AIB and ISOVAL on hydrolysis by chemical³ or microbial⁴ agents. The ratio of the two hydantoins in gasifier water⁵ was in fact the same as that observed for the AIB and ISOVAL in K/T sediments¹. Because hydrolytic conditions were used to obtain the amino acids from the sediments, the actual species in the sediments is unknown.

Hydantoins were formed in the gasifier condensate water by the complex Bucherer-Bergs reaction of ammonia, carbonate, hydrogen cyanide and various ketones and aldehydes that collected in the gas cooling/condensing system of the gasifier and not from a high-temperature reaction in the gasifier itself^{6,7}. In the

upper pyrolysis region of the gasifier, hot gases produced in the high-temperature gasification zone heat the coal and release large amounts of organic components, including many ketones (mainly acetone and 2-butanone). The ketones react in the condensed state with the dissolved gases, giving yields as high as 4 g of the two hydantoins per kg of coal.

Intrusion of magma into a coal bed would produce the intense heat needed for natural coal gasification to produce the hydantoin precursors. Ground water may have helped to produce steam, which could have effectively driven the gasification, and ground or surface water may also have provided a coolant for the gases, so that they could react to form the hydantoins. From the yields given above, geothermal gasification of a small coal seam of 1 km² by 10 m thickness and subsequent hydrolysis could have generated 6×10^7 kg of amino acids, enough to explain the abundance in K/T sediments. Further evidence for this process is the sedimentary deposit of char and coal particles in the Lattengebirge, Bavaria, K/T section⁸.

Whereas the K/T iridium anomaly requires a mantle type of volcanic activity or bolide, any kind of volcanic or magmatic intrusion into a coal field could have been responsible for the formation of the hydantoins. This may have been an event at the periphery of

the European deposition area, as suggested by Graup *et al.*⁸, or perhaps associated with the North American Laramide magmatic trend⁹, such as the Raton area in Colorado, where large areas of charred coal have been observed¹⁰. Because coal deposits as well as geothermal activity were plentiful from New Mexico to Canada, coal gasification may have occurred in several localities.

Zhao and Bada¹ dismissed the possibility of a volcanic origin for the abiotic amino acids on the grounds that high temperatures associated with volcanism would destroy them, and concluded that they were delivered by a massive extraterrestrial object (bolide) colliding with the Earth. The improbability of survival of amino acids in a bolide collision of the magnitude postulated, and the unlikely amount of diffusion that would be needed to have distributed the two amino acids away from the iridium-enriched K/T layer into the sediments above and below has been addressed by others^{11,12}.

Edwin S. Olson

Universal Fuel Development

Associates, Inc.,

223 Circle Hills Drive,

Grand Forks, North Dakota 58201, USA

1. Zhao, M. & Bada, J. L. *Nature* **339**, 463-465 (1989).
2. Olson, E. S., Diehl, J. W. & Miller, D. J. *Analyt. Chem.* **55**, 1111-1115 (1983).
3. Ware, E. *Chem. Rev.* **46**, 403-470 (1950).
4. Yamada, H., Takahashi, S., Kii, Y. & Kumagai, H. *J. ferment. Tech.* **56**, 484-491 (1978).
5. Willson, W. G., Hendrikson, J. G., Mann, M. D., Mayer, G. G. & Olson, E. S. *Proc. Techn. Utilization Low-Rank Coals* **1**, 265-299 (1984).
6. Olson, E. S., Worman, J. J. & Diehl, J. W. *ACS Div. Fuel Chem. Preprints* **30**(2), 288-290 (1985).
7. Diehl, J. W., Olson, E. S. & Worman, J. J. *Fuel* **64**, 1019-1021 (1985).
8. Graup, G., Spettel, B., Herm, D. & Weidich, K. F. *Earth planet. Sci. Lett.* **95**, 271-290 (1989).
9. Officer, C. B. *J. geol. Educ.* **38**, 402-416 (1990).
10. Crelling, J. C. & Dutcher, R. R. *Geol. Soc. Bull. Am.* **79**, 1375-1386 (1968).
11. Cronin, J. R. *Nature* **339**, 423-424 (1989).
12. Zahnle, K. & Grinspoon, D. *Nature* **348**, 157-160 (1990).

Script handedness

SIR — Dale Sanders suggests (*Nature* **355**, 483; 1992) that the direction of writing of a script depended originally on the medium in which it was written; that a naturally right-handed stone-carver would tend to chisel from right to left while a right-handed scribe using liquid ink would tend to write from left to right to avoid smudging. Consequently, I wonder whether there is a greater incidence of left-handedness in countries that use Arabic script than in others that use Roman or Devanagari, for example?

Barry Knight

Ancient Monuments Laboratory,

English Heritage,

23 Savile Row,

London W1X 1AB, UK

Triumph of the embryo?

Mark Ridley

The Meaning of Evolution: The Morphological Construction and Ideological Reconstruction of Darwin's Theory. By Robert J. Richards. *University of Chicago Press*: 1992. Pp. 205. \$19.95, £15.95.

"HARDLY any point gave me so much satisfaction when I was at work on the *Origin*, as the explanation of the wide difference in many classes between the embryo and the adult animal, and of the close resemblance of the embryos within the same class." The author is Charles Darwin, and he is reflecting on von Baer's law (or something like it): the law that the embryonic stages of two species, say dogs and humans, are more similar than are their adult forms.

Darwin explained the law by two processes. Natural selection would tend to alter the adult more than the embryo, which is relatively protected in many species; and variations are usually inherited at the same age in the offspring as they appear in the parent. Thus if natural selection favoured a variant in the adult, then only the adult, and not the embryonic, stage of the offspring would be modified. The embryos of different species would then be left looking similar, whereas the adults would diverge to look different. Darwin clearly states this theory in chapter 13 of the *Origin*, and clearly offers it as an explanation of von Baer's law: of the nine pages of the relevant section in my copy, seven are about von Baer's law and the final two pages consider an exceptional case and a possible extension of the principle to allow one to infer ancestral stages.

... to infer ancestral stages. Darwin's exact meaning in the one crucial sentence is not perfectly clear, but it is at least close to the idea of recapitulation (that is, that embryonic development recapitulates the species' ancestry). But whatever he meant there, he was undoubtedly interested in recapitulation and he discussed the idea on several occasions. It is from those scattered remarks that Robert Richards has built his case. Richards argues that Darwin did not follow von Baer's embryology, but was instead a full-blooded recapitulationist: Darwin believed in recapitulation and built a theory of evolution in which recapitulation "had to be true"; Darwin's evolutionary theory was embryological, and required that species inevitably evolved progressively, from

lower to higher forms, like an egg developing into an adult.

Richards suggests, moreover, that the idea of embryological recapitulation was the most important influence in Darwin's own thinking about evolution; nothing else "was quite as encompassing and significant for his general conception of

book). Richards is a leading historian of science, and professional Darwin scholars will have to study what he has to say. But after reading the book and re-reading Darwin's own remarks, I am unconvinced. Indeed, I am more than unconvinced: I think Richards' whole argument is dotty.

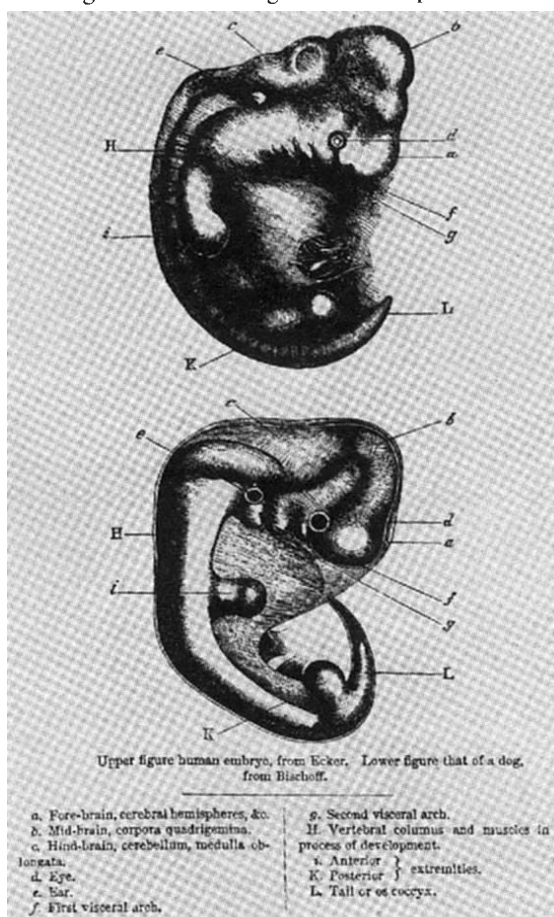
The first part of the book retraces the pre-darwinian history of recapitulation, and is not controversial; the trouble starts when Richards moves on to Darwin. The materials he has selected from Darwin do not (it seems to me) support the interpretation he places on them. The selection is unrepresentative and tendentious. And although some of the passages undoubtedly are about recapitulation, I do not detect in them the

enthusiasm that Richards sees. Other passages are clearly not about recapitulation at all: they look as if they are about von Baer's law. Thus, when Richards wishes to show that recapitulation was crucial in Darwin's first conception as well as in the mature theory, he turns to those obscure passages in the notebooks before Darwin read Malthus's *An Essay on the Principle of Population*.

Some of Darwin's ideas were indeed embryological at that time, but the simplest reading of history suggests (contrary to Richards' belief) that he threw them all overboard when he discovered natural selection in 1838. Then, in the "Sketch" of 1842, Darwin wrote "it is not true that [in individual development] one passes through the form of a lower group": Richards calls this remark "short-lived". In the "Essay" of 1844, Darwin was more sympathetic, and showed how natural selection could result in recapitulation: Richards concludes that this shows "the depth of Darwin's commitment to recapitulation". In the *Origin*, as we have seen, Darwin was concerned with von Baer's law and mentioned recapitulation at the end as an interesting hypothetical extension: Richards concentrates on the short

final passage. In later writings, Darwin sometimes took recapitulation more seriously again. Darwin's remarks about recapitulation thus range from dismissing it as an error to treating it conditionally, as an empirical possibility. They do not suggest that he was deeply committed to the process.

They suggest instead that, as earlier historians such as Michael Ghiselin have argued, Darwin had a general theory about developmental change in evolution; and that he applied it as the fluctuating empirical opinions of embryologists demanded. If recapitulation was



Embryo homologues. From Darwin's *Descent of Man*.

the evolutionary process" as "the principle of recapitulation and the embryological model". It would be no easy task to show that recapitulation (about which there are one or two vague sentences in *Origin* and a few more elsewhere) was a bigger influence on Darwin than, say, artificial selection (which gets a full chapter in the *Origin* and innumerable references), geographic variation (recurrently discussed in many books), classification (ditto, including four volumes on barnacles) or inheritance (a chapter or two in the *Origin*, years spent breeding domestic varieties, and a two-volume

wrong and von Baer right (as Richard Owen and Thomas Henry Huxley urged in the 1840s and 50s), then the application to von Baer was needed. If recapitulation really did happen after all (as Fritz Müller and Ernst Haeckel later argued), then that could be explained too. If there were exceptions, such as barnacles with their complex larval and simplified adult forms, Darwin could also reason about them: not, as Richards supposes, to show that they "did not refute the general principle of recapitulation", but as an application of a general theory to a test case.

There is no doubt that Darwin was interested in embryology. It provided important evidence for evolution and stimulated him to apply his theory in the manner that gave him "so much satisfaction". But nothing in Richards' book proves that Darwin believed recapitulation "has to be true" or that it was a necessary rather than a contingent "deductive consequence" of Darwin's general theory. □

Mark Ridley is in the Departments of Anthropology and Biology, Emory University, Atlanta, Georgia 30322, USA.

Investigating ancient enigmas

Robert Temple

Natural Knowledge in Preclassical Antiquity. By Mott T. Greene. Johns Hopkins University Press: 1992. Pp. 182. \$28.50, £18.

MOTT Greene is a brilliant and original investigator of ancient enigmas. This lucid and thoroughly accessible book contains seven essays exploring aspects of knowledge in antiquity up to the time of Plato. The first essay is a devastating attack on conventional notions of 'pre-history'. Greene tears to shreds the cosy assumptions that there can be such a "buffer zone" between "the biological ascent of hominids" and "the 'ascent to civilization' of the abstract 'mankind' of humanistic historical writing." He exposes the wishful thinking, the unwarranted assumptions and the modern myth of "the boundary between 'us' and 'them'" — "the myth that we are somehow special and different or even new".

The central importance of the book, however, is Greene's discovery that Hesiod's *Theogony*, a Greek mythological poem of the eighth century BC, preserves eyewitness descriptions of the eruptions of two ancient volcanoes — Thera (Santorini) circa 1470 BC and Mount Etna in Sicily (either the eruption coincident with Thera's or the eruption of 735 BC). As the author of a book on the history of geology, Greene knows this subject well. He is aware, for instance, that in the past 10,000 years there have been 5,564 identifiable eruptions of 1,343 volcanoes, 627 with specifiable dates. But the close survey of several volcanoes in the Mediterranean has allowed him unambiguously to match two of them to the two different eruptions described in *Theogony*.

The eruption of Thera is portrayed in the section of the poem known as the Titanomachy. Greene breaks down the poem's description into 15 successive stages and shows that these precisely

match the kind of eruption that took place at Thera, and supplements the work with much fascinating volcanic information, including results of a study in 1963 of the volcanic island of Surtsey. Green's case is strongly convincing, especially as "there is a complete one-to-one correspondence with no missing elements and . . . they are all in the correct order." This is an important discovery, showing that the violent Thera explosion was so traumatic that a detailed eyewitness account of great accuracy was preserved for seven centuries before being incorporated into Hesiod's mythological poem.

Greene also seems to have resolved, in a truly ingenious manner, the interminable debate over the identification of the sacred intoxicant *soma* in ancient India. It was *not* the fly-agaric mushroom. And he explains how Thales, credited with being the first scientist philosopher of Greece, actually managed the 'impossible' feat of diverting the Halys River to enable King Croesus's army to cross. As a hydraulic engineer, Thales understood the principles of meanders and diverted the river into an empty crescentic channel, a remnant of a former oxbow lake, merely by cutting through the silted-up entrances upstream and downstream. Now flowing in channels, the river became fordable without undue effort or time spent digging a fresh channel. So the diversion of the Halys was not a mere "mythical accomplishment".

Greene is a promising iconoclast whose investigative work should continue. He explains that the MacArthur Foundation made this book possible — let us hope more such results emerge from its largesse. □

Robert Temple is the author of *Conversations with Eternity*, which also investigates natural knowledge in antiquity, and may be contacted c/o David Higham Associates Ltd., 5-8 Lower John Street, Golden Square, London W1R 4HA, UK.

Celestial beams

R. J. Cohen

Astronomical Masers. By Moshe Elitzur. Kluwer: 1992. Pp. 351. Dfl.75, \$45, £26.

It is doubtful whether anyone could have predicted the existence of celestial masers. Yet the conditions for maser action, so difficult to produce in the laboratory, are easily achieved in interstellar space. Molecules warmed by an energy source, such as a young star, can be excited into a population inversion (so that more molecules are in high energy levels than in the ground state) over a wide range of conditions. The lines in the molecular spectra are then amplified to produce powerful beams of radiation detectable even in distant galaxies. Masers give us unique information about comets, circumstellar matter, star-forming clouds and active galactic nuclei, and also the intervening interstellar medium that is probed by the maser beam. With the development of radio-mapping techniques, masers have become an important astronomical tool, allowing us to study these regions on milliarcsecond angular scales.

Celestial masers were discovered by accident, and serendipity continued to play a large part in the development of the subject until quite recently. Part of the difficulty is inherent. Because maser radiation is strongly beamed, we cannot recover the source structure and physical conditions uniquely from our observations. But another part of the difficulty was simply historical timing. Masers were discovered too early. Proper understanding of the phenomenon had to await developments in radio interferometry and access to the millimetre and far-infrared wavebands. Now that these are all available, the picture is clearing, and new maser lines are at last being successfully predicted.

With the subject entering a new phase of maturity, Moshe Elitzur has chosen an ideal time to produce this book. It is aimed at "graduate students embarking upon research in astronomy" and is destined to become the standard introduction to the subject. Elitzur is particularly successful at relating the theory of masers to undergraduate physics. Not all graduates will have the necessary grasp of molecular spectra to be content with the short summary given here, but in other respects the book is exemplary. The author develops difficult concepts of radiative transfer, escape probability, pumping, saturation and beaming, clearly and thoroughly, adopting the didactic style of standard textbooks on electricity and magnetism. He carefully explains

the basic assumptions in what one feels at times is a crusade to purge the literature of half-truths. Some detailed proofs and alternative derivations are left as (well-hinted) exercises for the reader. The approach works well for maser theory, which occupies the first two-thirds of the book. This part is not always easy going, but the reward for the hard climb is a broad overview of the subject, from the theoreticians' perspective.

In the last third of the book, Elitzur reviews the observed masers and their use as astronomical tools. This material will not submit readily to textbook treatment, so here the author's style changes to that of a monograph or review. The coverage is broad and comprehensive, but necessarily more superficial than the treatment of maser theory. Researchers will want to consult the observational review articles that are cited, where they will also get a stronger taste of the excitement of current research and the glamour and spectacle of masers. Masers are not only intellectually challenging, they are also fun, and this is one ingredient that I missed in this otherwise clear and excellent introduction. The book is an essential purchase for workers in the field, and will also interest other astrophysicists and researchers in related areas of physics. □

R. J. Cohen is at the Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK.

Crab canon

F. A. Huntingford

Assessments and Decisions: A Study of Information Gathering by Hermit Crabs.

By R. W. Elwood and S. J. Neil. *Chapman and Hall*: 1991. Pp. 192. £29.95.

How does natural selection act on behavioural traits? Or more specifically, what are the effects of animals' behaviour on their fitness? In recent years (some might say not before time), behavioural ecologists have recognized that to answer these questions it is often necessary to understand the mechanisms that generate adaptive behavioural responses — in other words, to study what causes animals to behave as they do. But integrating studies of the causes and consequences of behaviour is not at all easy. In *Assessments and Decisions*, Elwood and Neil address this task, using data from an elegant body of experimental research conducted over the past decade or so on decisions made by hermit crabs as they seek out and fight over the shells in which they live.

Hermit crabs in general (and in particular the small species such as *Pagurus bernhardus*) are well suited for such studies. Their shell is a precious resource, the species and size of which can make the difference between life and death. Encounters between crabs and shells, and between crabs and crabs, can

easily be set up in the laboratory, and the responses that the crabs use to assess shells and opponents are clearly visible and easy to record. These encounters are illustrated by the many exquisite photographs in the book.

Elwood and Neil begin by defining and discussing the basic concepts. The term 'assessment' is mainly used to describe situations in which "some aspect of the environment is perceived and the information [is] used as a factor in decision making"; 'decision-making' is used to refer to "a process by which behavioural change is non-random". The authors next describe the behavioural ecology of hermit crabs and the biology of their study population of *P. bernhardus*. The scene is then set for the two main sections of the book in which motivational models are developed to explain the crabs' shell-related behaviour.

Elwood and Neil describe how crabs assess the quality of an unoccupied shell, presenting a single-factor motivational model of the behavioural mechanisms involved. They then provide an account of the shell fights, explaining in the context of games theory the role of shell value and opponent size in decision-making. The single-factor model for unoccupied shells is then expanded into a two-factor graphical model. Here, a crab's assessment of the value of the contested shell and of the probable costs of continuing a fight determine its choice of responses. The authors scrutinize how well their models explain the crab's behaviour and describe experimental tests of the models' predictions. They conclude by explaining how their models go beyond others, expressing the hope that this "will result in advances in our appreciation of how animals make assessments and decisions".

This is a very good and enjoyable book, and the experiments are described in a clear, concise and accessible way. Elwood and Neil address an important question of general interest to behavioural ecologists and "those interested in motivational analyses" (the intended audience of the book). Nevertheless, I have mixed views about whether the hope expressed at the end is realistic. Although the experiments described in the book contribute greatly to our understanding of how animals make decisions, I do not feel that the models (in the graphical form in which they are presented) add much to a simple verbal statement of the hypotheses they embody. This may well be just a personal quirk, though — it certainly does not detract much from the book's considerable value. □

F. A. Huntingford is in the Department of Zoology, University of Glasgow, Glasgow G12 8QQ, UK.



Warfare interest in radar technology and microwaves led directly to the development in 1951 of the first maser device by Charles H. Townes (middle) and colleagues at Columbia University. But the original idea also seems to have been independently conceived by Joseph Weber at the University of Maryland, and by Alexander M. Prokhorov (left) and Nikolai G. Basov (right) at the Lebedev Physics Institute in Moscow. The field grew rapidly and involved many of the people who later developed the laser. This fascinating and often controversial history is told in Jeff Hecht's *Laser Pioneers*, which contains 13 interviews with the researchers who made the breakthroughs. A revised edition is now published by Academic Press, price \$39.95.

Yeast roll

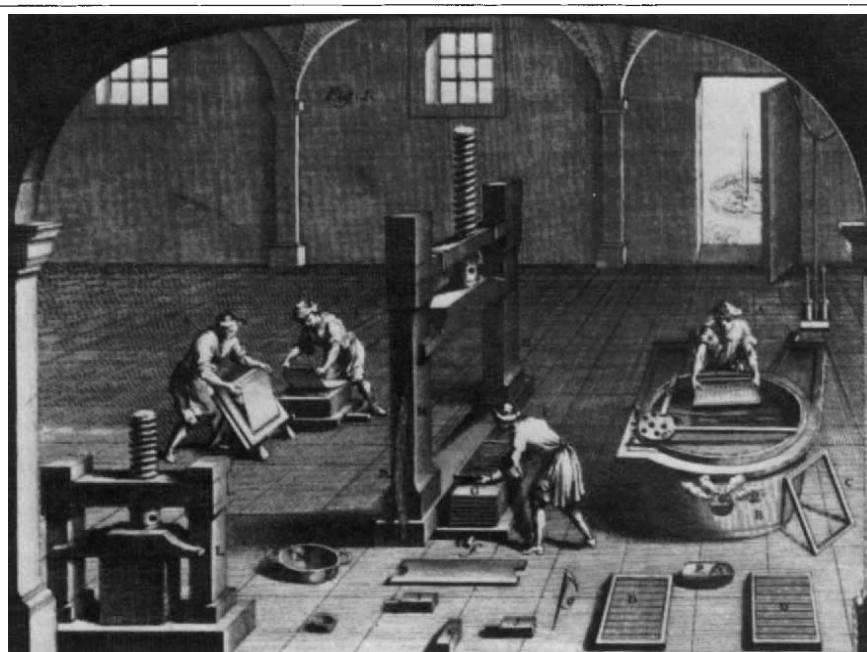
David Wilkie

The Molecular and Cellular Biology of the Yeast *Saccharomyces*. Volume 1: Genome Dynamics, Protein Synthesis and Energetics. Edited by James R. Broach, John R. Pringle and Elizabeth R. Jones. Cold Spring Harbor Laboratory Press: 1991. Pp. 826. \$97 (hbk), \$55 (pbk).

THE yeast *Saccharomyces* has come of age as a model of the eukaryotic cell. Not only is there close correspondence with higher organisms in the genetic control of fundamental processes such as cell-cycle progression, messenger RNA processing and the biogenesis of membrane systems, particularly those of the mitochondrion, but the main gap of differentiation has been breached by genetic engineering: the genes that specify the biochemistry of a liver cell can be transferred and expressed in the yeast cell. It is fitting that we now have this updated reference source giving a detailed account of the recent advances in the molecular biology of the organism.

The work provides a comprehensive and extensive coverage. It is wide ranging, each of the 11 chapters having been compiled by investigators from laboratories specializing in a particular area. It is possible, however, to group the various contributions into three broad categories, the first of which can be listed under the heading of the nuclear genome. The 16 chromosomes are catalogued according to size as determined by chromosome separation by pulsed-field gel electrophoresis, and procedures for constructing restriction maps are clearly laid out. Attention is given to *cis*-acting sequences, their role in chromosome function and their use in making artificial chromosomes. The mechanisms of mitotic and meiotic recombination, for many years the subject of argument and discussion among geneticists, are thoroughly described, as are the associated procedures for the integration of transforming DNA. Mutagenesis comes under scrutiny in a chapter devoted to chromosome damage and repair mechanisms, including excision repair, photoreactivation and recombinational repair.

A suitable title for the second category of topics would be 'cytoplasmic inclusions'. A chapter on DNA plasmids deals mostly with the composition, expression and maintenance in the yeast cell of the 2-micron circle. RNA viruses, mainly those of the killer type, are treated in a separate chapter. Although the route of infectivity of these double-stranded RNA particles by cell-to-cell



Paper was invented in China at about the time of the birth of Christ, but it was not until the twelfth century that the art of papermaking reached Europe. There, unlike in Asia, papers were made almost exclusively from old cloth rags. As paper consumption increased with the development of printing, there were chronic shortages of rags. In 1666, the English Parliament passed a law forbidding the use of cotton and linen for burying the dead. This engraving of a French paper mill around 1761 is taken from *Recycled Papers: The Essential Guide* by Claudia G. Thompson. Published by MIT Press, price \$40, £35.95 (hbk), \$25, £22.50 (pbk).

fusion distinguishes them from their animal counterparts, in other fundamental respects, such as replication and transcription processes, they are similar and form a model for the study of these viruses in higher organisms. Transposable elements are dealt with in a chapter on the family of Ty transposons, underlining their relatedness to one another and to those of other eukaryotes. Special reference is made to the organization and promoter activity of these 'long terminal repeats'. The chapter on cell membranes focuses on biophysical aspects of plasma and vacuolar membranes, giving an overview of transport systems followed by a detailed description of plasma membrane and vacuolar H^+ -ATPases. Plasma-membrane permease systems and ion channels are also included in a treatise that is very much in the realms of the specialist but nonetheless timely — the function of membranes is a central problem in biology.

The biogenesis of mitochondria, described here in depth, is the epitome of cellular complexity, requiring the integration of genetic information from organelle and nucleus. Remarkably, the complement of genes in the mitochondrion and their products (inner-membrane proteins and elements of the translation machinery) are more or less the same for all mitochondrial systems. The bulk of the genetic information

specifying mitochondrial components resides in the nucleus. How the translation products of the nuclear genes, formulated on cytoplasmic ribosomes, gain access to the organelle across its membranes is an intriguing problem ably dealt with in an erudite exposition of the possible mechanisms involved.

The third category of topics concerns biosynthetic systems. A chapter on the assembly of the yeast ribosome covers the processing of precursor RNAs, the proteins involved in constructing the subunits and the biogenesis of the mature ribosome with the involvement of the nucleolus. A final chapter on translational control stipulates the elements required for the initiation of protein synthesis and the components of the translation machinery.

The book concludes with an appendix of genetic and physical maps of *Saccharomyces cerevisiae* (although for the 'complete sequence of yeast chromosome III' one must read the 7 May issue of *Nature*).

The editors have put together a commendably detailed book. Extensive coverage of this kind is a boon to all those working with *Saccharomyces* and might even encourage those with only fringe interests to convert to the organism.

David Wilkie is in the Department of Biology, University College London, Gower Street, London WC1E 6BT, UK.

A compositional classification scheme for meteoritic chondrules

Derek W. G. Sears*, Lu Jie*, Paul H. Benoit*, John M. DeHart*† & Gary E. Lofgren†

* Cosmochemistry Group, University of Arkansas, Fayetteville, Arkansas 72701, USA

† SN-2, NASA Johnson Space Center, Houston, Texas 77058, USA

A taxonomic scheme is proposed for the main component of the primitive chondrite meteorites—the chondrules. The scheme provides insight into the variety of chondrules originally produced in the primordial solar nebula, and the effects on them of secondary processes on meteorite parent bodies. It may also contribute to understanding the origin of compositional diversity in the chondrites.

THE nine classes of meteorites known as the chondrites are the most primitive known rocks in the Solar System¹. The main constituents of these rocks are chondrules, roughly spherical aggregates of silicate minerals unlike anything found in lunar and terrestrial rocks^{2,3}. Chondrules have therefore attracted considerable attention throughout the history of meteorite studies. Not only are they crucial to understanding their host meteorites, but they also afford unique insights into processes that occurred in the dusty, turbulent and apparently energetic nebula in which the Sun and planets formed.

The compositional trends displayed by the nine chondrite groups reflect processes occurring in the pre-planetary nebula, and therefore may be relevant to planet formation, but the details are contentious. Some authors argue^{4,5} that the compositional trends reflect the proportion of chondrules to matrix, so that chondrule formation is the principal process affecting the composition of dust in the nebula. Other authors have argued^{6,7} that the chondrite groups reflect compositional trends caused by the separation of nebular gas and dust before chondrule formation. Part of the reason for the lack of agreement is the variety and complexity of chondrules and, in our opinion, the lack of a systematic approach to chondrule study.

Whatever the processes that were responsible for the formation of chondrules, they produced a wide variety of textures and compositions. To complicate matters further, once on their parent bodies the meteorites were altered, mainly by thermal metamorphism but also by the passage of aqueous fluids. A comprehensive classification scheme for chondrules is clearly needed, but none of those so far suggested^{3,8–11} has proved entirely satisfactory.

Existing schemes depend on texture^{3,8,9} or a combination of texture and mineral chemistry^{10,11}. Texture is important in deducing certain aspects of chondrule formation (such as post-formation cooling rates), but depends on many apparently random factors such as the type and abundance of nucleation sites. The schemes that combine mineral chemistry and texture have also proved helpful, but are confusing when applied to metamorphosed meteorites. They also concerned fairly homogeneous subsets of the chondrule population, so that arguments were inevitably based on an incomplete data base, and satisfactory discussions of mass balance, and the relevance of chondrule compositions to whole rock compositions, were impossible.

We now propose a compositionally based classification scheme that will help in systematizing the wealth of data available and in disentangling the effects of nebular and subsequent processes. The scheme builds on the strengths of earlier work, using fundamental properties of the chondrules, is applicable to >95% of the chondrules and is easily applied by anyone with an electron microprobe. Most important, the new scheme stresses the original diversity of the chondrules and the complex yet facile way in which chondrules respond to parent-body metamorphism. Our results suggest that arguments against an

important role for chondrules in determining the compositional trends of the chondrites were premature.

Chondrules: diverse and altered

The earliest workers described the chondrules as 'globules'¹², and in the mid-nineteenth century they were referred to 'drops of fiery rain'¹³. In the present century they were thought to be silicate liquids which had condensed in the primordial solar nebula¹⁴, or impact-produced melts¹⁵. None of these ideas is currently popular, and other suggestions abound¹⁶. There is agreement, however, that the chondrules have been flash-heated. Some chondrules are pure glass, whereas some have textures indicative of once having been wholly melted but subsequently crystallized into a variety of textures depending on composition, nucleation processes, temperature and cooling rate¹⁷. Others have textures which are 'igneous', but surprisingly coarse-grained, and some part of these may never have been totally melted¹⁸. Kieffer refers to the former as 'droplet' chondrules⁸, and Dodd and others refer to the latter as 'clastic' or 'lithic' chondrules. Dodd even suggests that the lithic chondrules are abraded fragments of previously existing igneous rocks³. Gooding and Keil developed a successful scheme for classifying chondrule textures⁹, sorting the chondrules into nonporphyritic (uniform fine-grained textures known as cryptocrystalline, radial pyroxene, barred olivine and granular) and porphyritic (coarse-grained with phenocrysts of olivine and pyroxene located in a mesostasis of highly variable composition).

Other workers have stressed the importance of mineral compositions and identified two contrasting extremes, the type IA chondrules with FeO-poor olivine and the type II chondrules in which the olivine is FeO-rich¹⁰. Type IB chondrules have also been identified in which olivine, slightly higher in FeO than is type IA, is enclosed in pyroxene¹¹.

A complication in studying the early history of any meteoritic material is the alteration processes they have suffered, most notably metamorphism. The 'petrologic types' of Van Schmus and Wood¹⁹ and Sears *et al.*²⁰, which rely on physical and mineralogical measurements, have been developed to assess the amount of metamorphism.

Cathodoluminescence and composition

We became interested in the diversity of chondrule compositions during studies of their cathodoluminescence (CL)^{21–23}. We first sorted the chondrules into those with bright CL (group A) and those with little or no CL (group B) and then further subdivided them according to the colour and intensity of the CL in the chondrule grains and the mesostasis that encloses the grains (Table 1). It became apparent that the CL properties were correlated with differences in composition. Thus we can take as our starting point for the classification scheme the composition of the chondrule mesostasis. As this is of ill-defined mineralogy (it is often a glass) its composition can best be expressed in

TABLE 1 Chondrule groups defined in terms of cathodoluminescence and mesostasis composition

	Mesostasis*	Chondrule grains
A1	yellow	red
A2	yellow	none/dull-red
A3	blue	red
A4	blue ($An \geq 50$)	none/dull-red
A5	blue ($An \leq 50$)	none
B1	none/dull-blue ($Qz \geq 50$)	none/dull-red
B2	dull-blue (Qz 30–50)	none/dull-red
B3	purple (Qz 15–30)	none

* An and Qz refer to normative anorthite and quartz (mol %) calculated from defocused electron-beam microprobe analysis.

terms of the 'CIPW norm', which refers to mineral abundances calculated from the bulk composition of the mesostasis. To a good approximation, chondrule mesostases are quartz-, albite- and anorthite-normative, and in the meteorites of lowest petrologic type the chondrules are uniformly spread across the quartz-anorthite-albite ternary (Fig. 1a). With increasing petrologic type, the proportion of quartz- and anorthite-rich chondrule mesostases decreases, and the data migrate to the albite-rich corner of the diagram. Clearly, parent-body metamorphism is homogenizing the mesostasis composition, and in the most metamorphosed chondrites (type 5 or 6) the quartz component is missing and the feldspar is fairly albitic ($An_{10}Or_6Ab_{84}$)²⁴.

As with the earlier work, the CL properties emphasize the importance of olivine composition. Unlike terrestrial and lunar olivines, some chondritic olivines contain high abundances of refractory elements, such as Ca, Ti and Al (ref. 25). The amount depends on the levels of parent-body metamorphism suffered by the meteorite, and on the individual chondrule. Figure 2 compares the CaO content of chondrule olivines with the FeO content for meteorites of four petrologic types. Like the mesostasis, chondrules in the type 3.0 chondrite contain olivines with a wide range of compositions, whereas chondrites of type ≥ 3.8 are fairly homogeneous. Chondrites at intermediate petrologic type may contain olivines with a greater range of CaO; presumably olivines in group B chondrules lose their Ca faster than the olivines in group A chondrules. This difference in the behaviour of olivines in the two groups of chondrule probably reflects the differences in Ca diffusion in Fe-rich and Fe-poor olivines. The plots for Al and Ti are almost identical to that of

Ca, although Ti lags slightly behind Ca and Al, and only olivines of type 4 or higher are uniform in Ti. There are a few chondrules in which olivine is absent but pyroxene behaves in a similar way, although at a slower rate because of lower diffusion rates.

Compositional classification scheme

We can use olivine and mesostasis compositions to divide the chondrules into eight groups, most of which have unique CL properties. Those whose mesostasis compositions are migrating down the quartz-albite axis constitute three 'B' groups; those moving along the feldspar axis constitute four 'A' groups. The compositions towards which both groups migrate correspond to those of group A5. Similarly, we can define fields on plots of the refractory element against iron to characterize the groups further. The A series, for example, shows systematic decreases in Ca and increases in Fe along the A1,2-A3-A4-A5 sequence (A1 and A2 are genetically related, so really A1 goes to A3, then A4 and then A5, while A2 goes to A4 then A5); Ca, in contrast, decreases along the B1-B2-B3-A5 series, although there is some overlap. These series are gradational and subdivision sometimes involves arbitrary values, but the discrimination plots shown in Figs 1 and 2 constitute a straightforward and workable means of defining the groups. In cases where olivine is absent, we can use the FeO content of the pyroxene and the FeO axis in Fig. 2. The A1 and A2 groups are closely related, showing many shared trends and similar bulk compositions involving major volatile element depletions and strong positive Eu anomalies²¹.

Inevitably this classification scheme preserves some of the features of earlier schemes. The Semarkona type IA chondrules of refs 10 and 11 are all members of the present group A1, and group A2 includes low-FeO porphyritic pyroxene chondrules that did not have a place in the previous mineral-chemical scheme. Group B1 includes the type II chondrules of previous workers. To a good approximation, the A groups consist of the 'droplet' chondrules, whereas the B groups constitute the 'clastic' chondrules. Unlike earlier schemes, the scheme proposed here does not depend on texture, nor is it restricted to a fairly homogeneous subset of chondrules. The textural terms of Gooding and Keil⁹ can be used with the present group definitions, describing for example a group A1 barred olivine chondrule or a group A5 porphyritic olivine-pyroxene chondrule.

With two exceptions, we can use the CL properties in Table 2 to determine the relative proportions of the chondrule groups in other chondrites without the need for analytical data. The

FIG. 1 Ternary plots for the calculated major mineral abundances in the chondrule mesostases of the Semarkona, Krymka, Allan Hills A77214, and Dhajala meteorites and a group of 'equilibrated' (essentially type 4) chondrites. The symbols for the minerals are: Q, quartz; Ab, albite; An, anorthite. The degree of parent-body metamorphism is reflected in the 'petrologic type' 3.0–4, which is assigned on the basis of physical, mineralogical and petrological criteria^{19,20}. The chondrules in the virtually unmetamorphosed Semarkona have a wide range of composition, but this decreases with increasing petrologic type as the data migrate to the albite corner of the diagram. The chondrule group assignments are purely on the basis of cathodoluminescence and are shown for comparison with the compositional definitions which are shown in the sixth ternary. Tie-lines connect data for chondrules in which the mesostases have outer zones and interiors with differing CL, and whose CL classification is sometimes difficult. This mainly applies to Allan Hills A77214. The data are from the present study except those for the equilibrated meteorites which are from ref. 35.

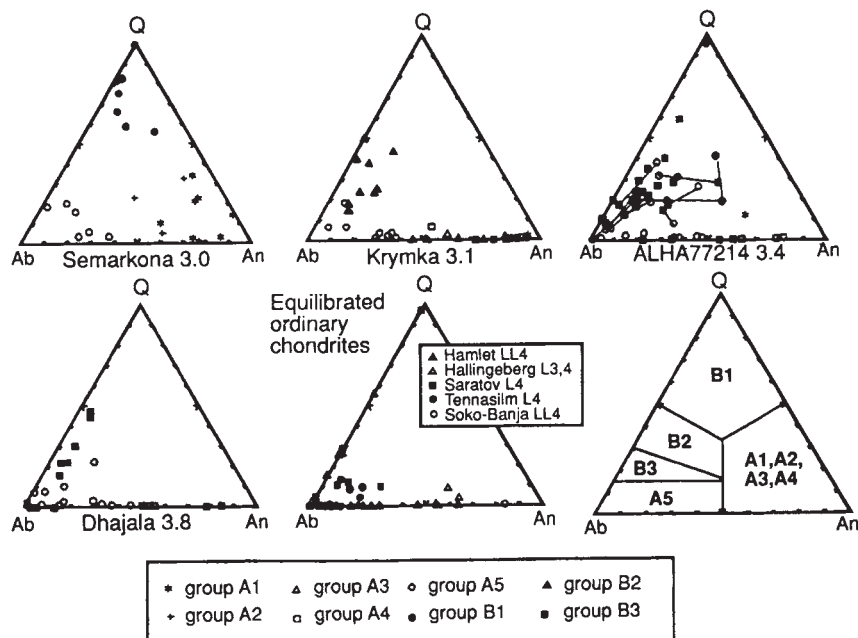


TABLE 2 Sections used here with frequency of chondrule groups as a function of petrologic type

Meteorite	Section number source*	Petrologic type	Number of chondrules	CL groups							
				A1	A2	A3	A4	A5	B1	B2	B3
Semarkona	USNM 1805	3.0	76	10.5	25.0	0.0	0.0	5.0	56.9	0.0	2.6
Bishunpur	BM 80399	3.1	98	7.1	0.0	12.2	12.8	12.8	0.0	51.0	1.0
Krymka	USNM L729-7	3.1	134	5.2	0.0	20.9	17.8	17.8	0.0	29.9	1.5
Chainpur	BM 191586	3.4	90	0.0	0.0	10.0	11.7	27.2	0.0	38.8	6.7
ALHA77214	MWG PTS#8	3.5	124	1.6	0.0	2.4	15.5	36.1	1.6	0.0	41.9
Hedjaz	BM 192513	3.7	69	0.0	0.0	0.0	21.3	49.7	0.0	0.0	29.0
Dhajala	PRL	3.8	62	0.0	0.0	0.0	18.4	42.9	0.0	0.0	38.7
Bremorvörde	BM 33910	3.9	60	0.0	0.0	5.0	19.0	44.3	0.0	0.0	31.7
Barwell	BM 1966	5	63	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0

Data obtained by assigning all the chondrules in these sections to the chondrule groups using their CL. Frequencies are per cent by numbers. Groups A4 and A5 cannot readily be distinguished by CL and the above statistics assume the following (based on the four analysed chondrites in bold type): **Semarkona**, 100% A5; **Krymka**, Bishunpur, 50% A5; Chainpur, **ALHA77214**, Hedjaz, **Dhajala**, 70% A5. We assume Barwell chondrules are entirely A5 because many studies show that the mesostases of equilibrated chondrites are sodic feldspar²⁴.

* USNM: R. Clark and T. Thomas, National Museum of Natural History, Smithsonian Institution (Semarkona polished thick section made at University of California, Los Angeles). BM: R. Hutchison, British Museum (Natural History). MWG: Meteorite Working Group of the NSF/NASA. PRL: D. Lal, Physical Research Laboratory, Ahmedabad (polished thin section made at Johnson Space Center).

exceptions concern the distinctions between the A4 and A5 chondrules and between the B1 and B2 chondrules. For the moment, we assume the B1 chondrules are unique to Semarkona. The A4 and A5 chondrules have similar CL properties and we have assumed an A4:A5 ratio for chondrules in the unanalysed chondrites based on analysed chondrites of similar petrologic type. We find that the proportion of group A to group B chondrules covers a small range within which the ratio of group A to group B chondrules increases steadily with petrologic type. Presumably the group A to group B ratio for the Semarkona chondrite is an original, pre-metamorphic quantity characteristic of all ordinary chondrites. On closer inspection, it seems that with increasing petrologic type, chondrules of group A1,2 disappear to be replaced by A3 and then A4, which ultimately merge into group A5. Similarly, group B1 evolves into group B2, then group B3 and ultimately A5. Evolution along the A series is much more rapid than evolution along the B series; that is, greater levels of metamorphism are required to homogenize the B series. This is shown schematically in Fig. 3.

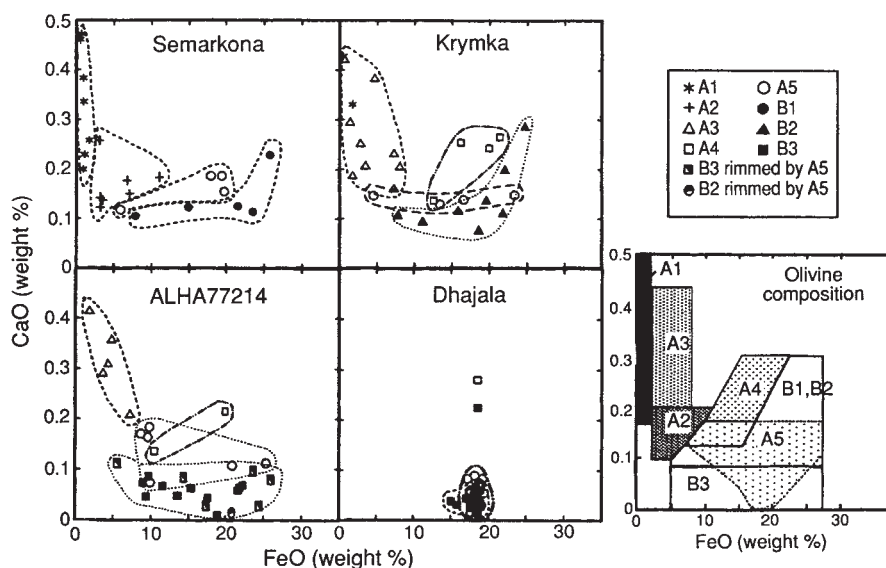
Deciphering meteorite history

The groups and the relationships described in Fig. 3 have many implications. The chondrule-forming process apparently produced four types of chondrule, of which two are strongly related, one of which resembles the chondrules in metamorphosed

meteorites in the bulk compositions of its major components. (Note, however, that mineral compositions in the A5 chondrules become more homogeneous with increasing petrologic type, showing a range of 5–25 wt% FeO in the type 3.0, 3.1 and 3.4 chondrites but only a 17–20 wt% range in the type 3.8.) Metamorphism increased the diversity of the chondrule population, but in a complicated way, as some chondrules respond to metamorphism faster than others. Individual chondrules may actually increase in mineral heterogeneity during metamorphism, because the smaller grains react faster than larger grains and secondary pyroxene is formed as the quartz component of the mesostasis is destroyed by reaction with olivine. The properties of the chondrules in a given chondrite therefore depend on the variety originally present, the range in their responsiveness to metamorphism and the metamorphism suffered by the meteorite as a whole. There is no reason to assume, because chondrules vary in their internal heterogeneity, that certain chondrules were metamorphosed before emplacement in the meteorite^{26,27}.

It has been suggested that chondrules as a whole do not show a relationship between calculated liquidus temperature and Na content, and that chondrules therefore did not suffer volatile loss during formation⁶. The group A1,A2 does show the expected trend, and group B1 is clearly genetically unrelated to group A1,2 (ref. 28), so it is possible that volatiles were lost during

FIG. 2 Compositions of chondrule olivines, the major silicate mineral in chondrules. Again chondrules in the least-metamorphosed meteorites have a wide range of compositions which disappears with increasing petrologic type. B2/A5 and B3/A5 refer to chondrules whose mesostases are zoned in their CL and compositional properties, B2 or B3 in the central regions and A5 in the outer regions. Aluminium and Ti show very similar patterns to Ca, the only important difference being that the olivines in A4,5 chondrules in Dhajala are significantly higher in Ti than the olivines in B3 chondrules in Dhajala. The volatile elements Mn, Cr and P behave very differently from the refractory elements and from each other, but the chondrule groups are still resolvable, albeit with some overlap. Assignments to groups has also been made on the basis of CL for comparison with the compositional definitions. Each chondrule can uniquely be assigned to a group using the composition of the two main chondrule components and compositional fields shown here and in Fig. 1. In a few chondrules in which olivine is absent, pyroxene composition can be used.



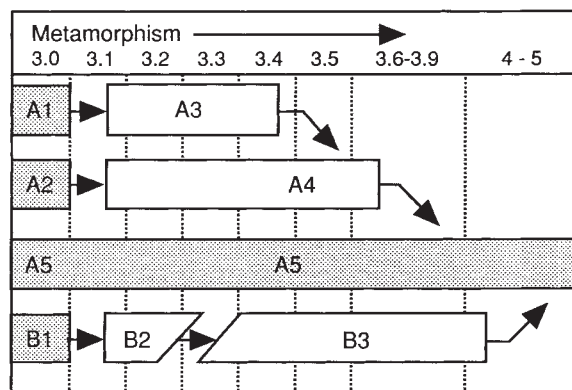


FIG. 3 Schematic representation of the proposed metamorphic relationship between the chondrule groups. Group A1, and the genetically related group A2, group A5 and group B1 are 'primary' (pre-metamorphic) and compositionally distinct. Metamorphism causes chondrules in the A and B series to undergo major compositional changes in their olivine and mesostasis and ultimately to converge on group A5 (Figs 1 and 2). Homogenization is easier, however, for group A, and considerable parent-body metamorphism is necessary to homogenize group B chondrules. An estimate of the level of parent-body metamorphism, as reflected in the petrologic type, is indicated by the type alongside each step. The mean olivine and mesostasis compositions of group A5 do not change during metamorphism, although the degree of composition heterogeneity of these phases decreases.

chondrule formation. It is also possible, however, that volatile content was a factor controlling liquidus temperature.

Processes in the early Solar System

The similarity of some compositional trends in the chondrites and the planets has often led to the suggestion that similar physical and chemical processes may have been responsible²⁹; that is, that the composition of the planets was controlled by the composition of their 'building blocks'. Both the planets and the chondrite groups show variations in their redox states, lithophile element abundance patterns, oxygen isotope systematics and siderophile element abundances. The cause of the compositional differences between the nine chondrite classes is controversial, the two main viewpoints being (1) that they reflect a nebula-wide process so that compositions are related to solar distance and the composition of the chondrules depends on their precursor^{1,7}, and (2) that they reflect the proportion of chondrules and matrix^{4,5,7,30}. There is an implicit assumption in the second model that chondrules are isochemical, but this is now known to be untrue.

As chondrules constitute the main fraction of the silicate portion of these meteorites, it is possible that the ratio of group A to group B chondrules in the original accreted material played an important part in determining the bulk compositions of the chondrites and thereby their class. The differences between group A and group B chondrules involve two, and perhaps three, of the relevant chemical trends. Group A chondrules are more reduced, have higher abundances of refractory lithophiles than group B chondrules and probably have different proportions of oxygen isotopes. Because of their small size and friability³¹, we have not yet obtained oxygen isotope data for group

A chondrules, but data for the Allende carbonaceous chondrite are at least suggestive of a systematic difference^{32,33}, and there seems to be a size dependency in the oxygen isotope for chondrules from the Dhajala meteorite³⁴.

In short, it seems possible that the differences between the chondrite classes reflect the proportion of group A to group B chondrules, rather than the proportion of chondrules to matrix. We think that the existence of such diverse groups of chondrules in the same rock, whose bulk composition is remarkably close to cosmic, may indicate that the chondrule groups were formed by the same processes acting with differing degrees of intensity on the same or very similar precursor material. Other arguments that the group A chondrules obtained their unusual compositions during chondrule formation have been summarized elsewhere²¹. It is therefore possible that the compositional variations in the meteorites, and planets, reflect differences in the intensity of chondrule-forming process at different points of the Solar System, rather than nebula-wide chemical processes.

Conclusions

We suggest that chondrules should not be classified by texture or the composition of a single phase, or by some mixture of the two, but in a comprehensive, systematic manner using the composition of the two main chondrule components. We believe that this will focus attention on the most meaningful meteorites, stress the range of materials present and provide a means of sorting them out by both pre-metamorphic composition and metamorphic alteration. We also argue that the recognition of these chondrule groups, and their various abundances in different types, has interesting implications for early Solar System studies. □

Received 18 September 1991; accepted 13 April 1992.

- Sears, D. W. G. *Vistas Astr.* **32**, 1-21 (1988).
- King, E. A. (ed.) *Chondrules and Their Origins* (Lunar and Planetary Institute, Houston, 1983).
- Dodd, R. T. *Meteorites. A Petrologic-Chemical Synthesis* (Cambridge Univ. Press, 1981).
- Wood, J. A. in *Protostars and Planets II* (eds Black, D. C. & Matthews, M. S.) 682-702 (Univ. of Arizona Press, Tucson, 1985).
- Larimer, J. W. & Anders, E. *Geochim. cosmochim. Acta* **31**, 1239-1270 (1967).
- Hewins, R. H. *Geochim. cosmochim. Acta* **55**, 935-942 (1991).
- Wai, C. M. & Wasson, J. T. *Earth planet. Sci. Lett.* **36**, 1-13 (1977).
- Kieffer, S. W. *Science* **189**, 333-340 (1975).
- Gooding, J. L. & Keil, K. *Meteoritics* **16**, 17-43 (1981).
- McSween, H. Y. *Geochim. cosmochim. Acta* **41**, 477-491 (1977).
- Scott, E. R. D. & Taylor, G. J. *J. geophys. Res.* **88**, B275-286 (1983).
- Howard, E. C. *Phil. Trans. R. Soc.* **51**, 168-212 (1802).
- Sorby, H. C. *Nature* **15**, 495-498 (1877).
- Wood, J. A. *Nature* **197**, 127-130 (1962).
- Urey, H. C. & Craig, H. *Geochim. cosmochim. Acta* **4**, 36-82 (1953).
- Grossman, J. N. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 680-696 (Univ. of Arizona Press, Tucson, 1988).
- Hewins, R. H. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 660-679 (Univ. of Arizona Press, Tucson, 1988).
- Nagahara, H. *Nature* **292**, 135-136 (1981).
- Van Schmus, W. R. & Wood, J. A. *Geochim. cosmochim. Acta* **31**, 747-765 (1967).

- Sears, D. W. G. *Nature* **287**, 791-795 (1980).
- Lu Jie *et al. Meteoritics* **26**, 367 (1992).
- Sears, D. W. G. *et al. in Spectroscopic Characterization of Minerals and Their Surfaces* (eds Coyne, L. M., McKeever, S. W. S. & Blake, D. F.) 190-222 (American Chemical Society, New York, 1989).
- DeHart, J. M. & Sears, D. W. G. *Lunar planet. Sci.* **XVII**, 160-161 (Lunar and Planetary Institute, Houston, 1989).
- Van Schmus, W. R. & Ribbe, P. H. *Geochim. cosmochim. Acta* **32**, 1327-1342 (1968).
- Steele, I. M. *Geochim. cosmochim. Acta* **50**, 1379-1395 (1986).
- Dodd, R. T. *Contrib. Miner. Petrol.* **31**, 201-227 (1971).
- Scott, E. R. D. *Smithsonian Contrib. Earth Sci.* **26**, 73-94 (1984).
- Sears, D. W. G., Lu Jie & Benoit, P. H. *Meteoritics* **26**, 393-394 (1992).
- Lewis, J. S. *Icarus* **16**, 241-252 (1972).
- Wood, J. A. *Lunar planet. Sci.* **XVIII**, 1100-1101 (Lunar and Planetary Institute, Houston, 1987).
- Sears, D. W. G. & DeHart, J. M. *Meteoritics* **24**, 325 (1989).
- Clayton, R. N. *et al. in Chondrules and Their Origins* (ed. King, E. A.) 37-43 (Lunar and Planetary Institute, Houston, 1983).
- McSween, H. Y. *Meteoritics* **20**, 523-540 (1985).
- Clayton, R. N. *et al. Geochim. cosmochim. Acta* **55**, 2317-2337 (1991).
- Gooding, J. L. thesis, Univ. of New Mexico (1979).

ACKNOWLEDGEMENTS. We thank the donors of the meteorites used here (Table 2), and M. Prinz, J. Wood and J. Grossman for discussions or reviews, W. Manger for the use of his cathodoluminescence microscope and NASA for funding.

Effect of polymorphism of an MHC-linked transporter on the peptides assembled in a class I molecule

Simon J. Powis, Edward V. Deverson, W. John Coadwell, Antonio Ciruela[†],
Neville S. Huskisson^{*}, Howard Smith, Geoffrey W. Butcher & Jonathan C. Howard

Department of Immunology and ^{*}Microchemical Facility, Agricultural and Food Research Council Institute of Animal Physiology and Genetics Research, Babraham, Cambridge, CB2 4AT, UK

Short antigenic peptides bound in the groove of class I major histocompatibility complex molecules enable T cells to detect intracellular pathogens. It has been assumed that structural features of the class I molecule alone select which peptides are bound. It is now demonstrated that a complex polymorphism in one of the major histocompatibility complex-encoded putative peptide-transporter genes is associated with an altered spectrum of bound peptides.

CLASS I major histocompatibility complex (MHC) molecules expressed at the cell surface are composed of a polymorphic heavy chain non-covalently associated with β_2 -microglobulin, and one of a range of short peptides, usually 8–9 amino acids long, derived from proteins synthesized in the cell^{1–7}. Assembly of all three components is required for normal surface class I expression, and the bound peptides contribute to the antigenic determinants recognized by T-cell receptors^{8–12}. We have observed a complex change in the expression of a rat class I MHC molecule, RT1.A^a, which occurred when the class I locus was associated by recombination with different alleles at the adjacent class II region of the MHC^{13,14}. Using conventional genetics we established that the new differential locus in the class II region, which we named *cim* (class I modifier), existed in two allelic forms, with one (*cim*^a) dominant over the other (*cim*^b) in F₁ hybrids^{15,16}. The main phenotypic change caused by *cim* allelism was a striking modification in the antigenicity of RT1.A^a as seen by T cells. Two T-cell-defined antigenic forms of RT1.A^a were expressed: A^{a+} in the presence of homozygous or heterozygous *cim*^a, and A^{a-} in the presence of homozygous *cim*^b. It was also observed that in the presence of the recessive *cim*^b allele alone RT1.A^a molecules were retained in the endoplasmic reticulum or cis-Golgi for an unusually long time^{15,17,18}. These two properties suggested that the *cim* product had some role in the assembly or loading of peptides into the nascent RT1.A^a molecule. *Cim* was subsequently mapped to a segment of ~250 kilobases of the rat MHC homologous to the *Pb-Ob* interval in the mouse *H-2*, or the corresponding *DP-DO* interval in the human *HLA* complex¹⁶. As part of the cloning strategy for *cim* we sought new genes in the *cim* region of the MHC. We and others recently reported the discovery of two genes in the *cim* region (named *mtp1* and *mtp2* in the rat), members of the ATP-binding cassette (ABC) transporter gene family^{19–22}. By transfection of various class I-deficient cell lines, including the mouse RMA-S and the human .134 and BM36.1 cell lines, both of these genes have been shown to be required for normal class I assembly and antigen presentation, findings consistent with the assembly of their protein products as heterodimeric complexes^{23–27}. These molecules, which are located in the endoplasmic reticulum membrane (J. Trowsdale, personal communication), are evidently candidates for the hypothetical peptide translocation mechanism

required to deliver peptides of cytosolic origin to the endoplasmic reticulum for class I assembly²⁸. So far no evidence has been forthcoming as to what constraints might exist in the selection of peptides for transport. We now show that sequence polymorphism in *mtp2* is responsible for the *cim* phenomenon, and that the gain and loss antigenic changes in RT1.A^a resulting from *mtp2* polymorphism are due to the loading of the RT1.A^a molecule with distinct populations of peptides. Thus while polymorphism localized in side-chain binding pockets of class I molecules does alter the requirements for peptide binding^{29,30}, it alone is not a sufficient explanation for the specificity of peptide uptake by class I molecules³¹. Furthermore, through their differential effect on peptide uptake, polymorphic transporter proteins may exercise specific influences on the development of the T-cell repertoire, and are consequently potential susceptibility factors in infectious and autoimmune disease.

Cim is the *mtp2* transporter chain

We first established an *in vitro* rat cell line suitable for the identification of *cim* by transfection. The rat T-cell lymphoma line C58-NTD (ref. 32) carries the RT1^a MHC haplotype and should therefore possess the recessive *cim*^b allele¹⁶. The C58-3.3/1 cell line was created by transfection of a cDNA encoding RT1.A^a (ref. 33). As expected, C58-3.3/1 cells expressed the transfected RT1.A^a product in the *cim*^b-associated A^{a-} form (Fig. 1a–d; and ref. 18). Our previous genetic studies showed that the *cim*^a allele is dominant over *cim*^b (refs 15 and 16). Introduction into C58-3.3/1 cells of a complementary DNA representing *cim*^a should therefore cause the expression of RT1.A^a in the A^{a+} form, which would be detectable with cytotoxic T lymphocytes (CTL) of appropriate specificity. Full-length cDNA clones encoding *mtp1*^a and *mtp2*^a, isolated from a DA, RT1^a library^{19,23}, were transfected into C58-3.3/1 cells. Transfection of *mtp1*^a resulted in no apparent alteration of phenotype (data not shown). When *mtp2*^a alone was transfected into C58-3.3/1, however, A^{a+} molecules were generated, as shown by lysis with A^{a+}-specific CTL (Fig. 1f). (Unexpectedly, A^{a-} was still expressed, because the *mtp2*^a transfectants were still lysed by anti-A^{a-} CTL as shown in Fig. 1e. The persistent expression of A^{a-} in these cells is discussed later.) Furthermore, in the C58-3.3/1.*mtp2*^a clones, RT1.A^a showed the pulse-chase phenotype of rapid glycan maturation during transit to the cell surface characteristic of the *cim*^a-associated A^{a+} form^{17,18}, in contrast to the slower processing observed in C58-3.3/1 cells

[†] Present address: Biological Laboratory, University of Kent, Canterbury CT2 7NJ, UK.

before *mtp2^a* transfection (Fig. 2). Taken together, these results indicate that the transporter gene *mtp2* is *cim*, with the implication that *mtp2* should show sequence variation responsible for the *cim* phenomenon.

Sequence polymorphism of *mtp2*

We previously identified the *cim* alleles of several rat haplotypes by their ability to generate the A⁺ phenotype in heterozygotes with a *cim^b* strain¹⁶. By this analysis the *RT1* haplotypes *a*, *d*, *f*, *l* and *o* were *cim^a*, and *b*, *c*, *k*, *n* and *u* were *cim^b*. If *mtp2* were indeed the molecular basis of *cim*, then *mtp2* alleles of different haplotypes should show polymorphic sequence variation correlating with their known *cim* type. We therefore generated the full-length sequence of polymerase chain reaction (PCR)-amplified *mtp2* cDNA from two *cim^b* haplotypes, *RT1^c* and *RT1^u*, and from one further *cim^a* haplotype, *RT1^l*. Figure 3 presents the amino-acid sequences of these four *mtp2* alleles using the published sequence of *mtp2^a* as a consensus²³. Although drawn from four unrelated *RT1* haplotypes, the *mtp2* sequences fell into two distinct groups correlating with *cim* type. In a total of 29 positions at which variation was recorded, 25 substitutions correlated with *cim* type. Figure 4 shows that these substitutions were situated predominantly in the N-terminal two-thirds of the sequence, which contains the hydrophobic, presumed membrane-spanning loops, whereas there were only 2 *cim*-correlated substitutions in the last 240 amino acids corresponding to the ATP-binding domain. Eighteen of the 25 *cim*-correlated substitutions were non-conservative, 11 of these involved charged amino acids, resulting in a net change in charge of 5. Six of the remaining *cim*-correlated changes were conservative hydrophobic exchanges involving valine. Several substitutions were clustered in groups of two or three adjacent to potential membrane-spanning regions. Figure 3 includes an alignment of the human *mtp2* homologue, PSF2 (ref. 34). The sequenced allele of PSF2 (which corresponds to the human *RING11B* allele³⁵) follows one or other of the *mtp2* alleles at 18 out of 25 of the polymorphic positions with a marked preference (12 out of 18) for the *mtp2^u*/*mtp2^c* (*cim^b*) residues. Certain of the clustered allelic substitutions in *mtp2^u* and *mtp2^c*, such as the

G-TL (single-letter amino-acid code) motif at 165–168 and the TM motif at 217–218, are preserved intact. It is also noticeable that the sequence in the region between the G-TL motif and the R–PF motif at 262–266 is strikingly conserved in this distant interspecific comparison.

Elution of peptides from RT1.A^a

Finally, we sought to explain the marked antigenic difference between the A⁺ and A[−] forms of RT1.A^a in the light of *mtp2* polymorphism. The difference is more easily detected by T cells than by serological methods¹⁵, a result consistent with the presence of different peptides associated with the two forms. Furthermore, as the antigenic difference was large enough to stimulate primary *in vitro* CTL responses, it seemed likely that relatively extensive changes in the peptide composition of the A⁺ and A[−] molecules were involved, and such indeed proved to be the case. Figure 5a–d shows ³H-Tyr/³H-Leu internally labelled peptides acid-eluted from RT1.A^a molecules immunoprecipitated from concanavalin A (con A) blasts of A⁺ (PVG.R19; Fig. 5a) and A[−] (PVG.R1 and PVG.R8; Fig. 5b and c) expressing rat strains, and separated on reverse-phase high-performance liquid chromatography. Apart from a system of late-eluting peaks common to all three profiles (Fig. 5a–c, small filled arrows) the A⁺ and A[−] profiles were remarkably and reproducibly distinct. In particular, the A⁺ profile was distinguished by a broad peak of material eluting early in the acetonitrile gradient (Fig. 5a, broad arrowhead); the two different A[−] profiles were characterized by more late-eluting material, and in particular by two peaks eluting around fraction 90 (Fig. 5b and c, open arrows). That the shift of RT1.A^a-associated peptides towards the hydrophilic end of the acetonitrile gradient in A⁺-expressing strains was due to the activity of *mtp2^a* was shown directly by analysing ³⁵S-Met internally labelled peptides eluted from RT1.A^a immunoprecipitated from C58-3.3/1.*mtp2^a* (Fig. 5e) and C58-3.3/1 (Fig. 5f) cells (as described in Figs 1 and 2). Again the presence of *mtp2^a* led to the recovery of a distinctive group of early eluting peaks in the acetonitrile gradient (Fig. 5e, broad arrowhead). Similar results were obtained with bulk and cloned transfectants using peptides internally labelled with ³H-Leu. We

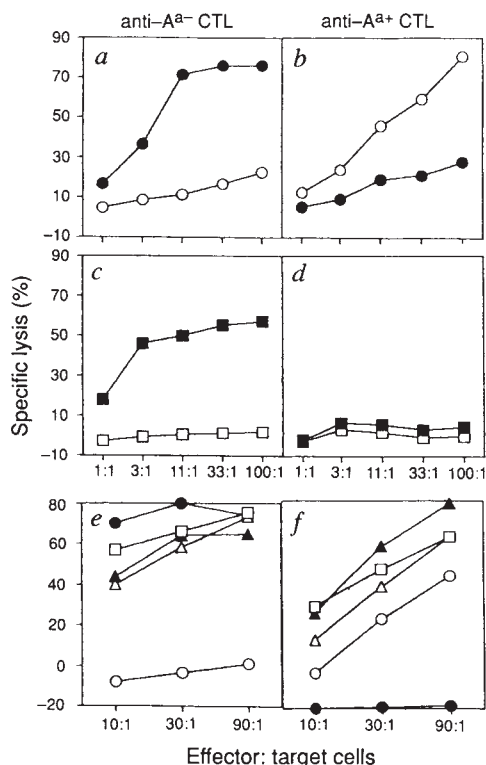


FIG. 1 C58-3.3/1.*mtp2^a* cloned cells are lysed by CTL specific for both the A⁺ and A[−] alloantigenic forms of RT1.A^a. *a*, Anti-A[−]-specific CTL (see Methods) lyse A[−]-expressing target lymphoblasts (●, PVG.R1), but not A⁺-expressing targets (○, PVG.R19). *b*, Reciprocally, anti-A⁺ CTL lyse A⁺ targets (○, PVG.R19) but not A[−] targets (●, PVG.R1). *c* and *d*, RT1.A^a-expressing C58-3.3/1 cells (■) are killed only by anti-A[−] CTL, but neither CTL population kills non-RT1.A^a-expressing untransfected control C58-NTD cells (□). *e* and *f*, C58-3.3/1.*mtp2^a* clones .A10 (▲), .B5 (□) and .H1 (△) are killed by both anti-A[−] (*e*) and anti-A⁺ (*f*) CTL; control cells are A[−] targets (●, PVG.R1) and A⁺ targets (○, PVG.R19).

METHODS. CTL specific for the A⁺ and A[−] alloantigenic forms of RT1.A^a were raised as described previously using the MHC-recombinant rat combinations PVG.R1, anti-PVG.R19, and PVG.R19 anti-PVG.R1 respectively¹⁸. Target conA-activated lymphocytes and cell lines were labelled with sodium [⁵¹Cr]chromate and incubated with various numbers of effector cells for 5 h before measurement of ⁵¹Cr release as described^{15,18}. The ordinate indicates per cent specific lysis, the abscissa indicates effector-to-target cell ratios. C58-3.3/1 cells were transfected with the full-length *mtp2^a* cDNA clone 441-11 (ref. 23) in the pHβApr-1-*neo* expression vector⁴⁴ by electroporation and subsequent selection in 1.0 mg ml^{−1} G418 (Gibco-BRL) as described²³. Clones .A10, .B5 and .H1 were isolated from the primary transfection by limiting dilution cloning in the same selection medium.

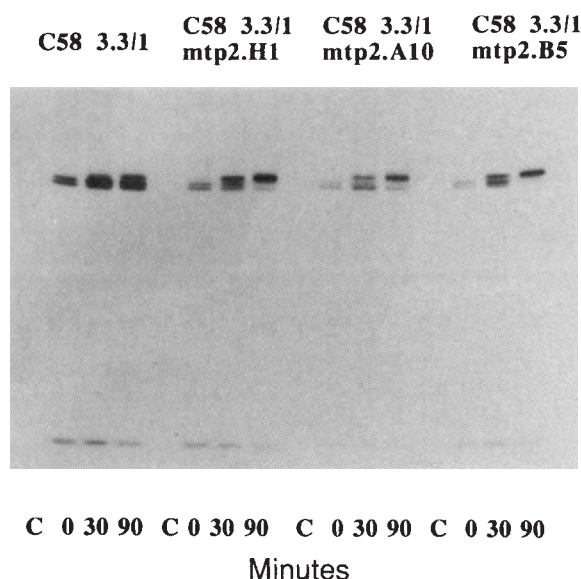


FIG. 2 Pulse-chase analysis of RT1.A^a in C58-3.3/1 and C58-3.3/1.mtp2^a cells (clones .A10, .B5 and .H1). Rapid intracellular transit of the class I molecule to the cell surface occurs in the *mtp2*^a-transfected cells, as indicated by the increase in *M_r* of the heavy chain during oligosaccharide side-chain processing. In C58-3.3/1 cells processing is far less rapid, indicating retention within the endoplasmic reticulum or *cis*-Golgi. Control tracks (C) indicate immunoprecipitation in the absence of anti-class I monoclonal antibody.

METHODS. Pulse-chase and immunoprecipitation of RT1.A^a were done as previously described¹⁸ using a 15-min pulse with ³⁵S-methionine (ICN-Flow) and immunoprecipitation with the RT1.A^a-specific monoclonal antibody AFRC MAC30 (ref. 45) coupled to Sepharose (Pharmacia). Samples were analysed by SDS-PAGE on 13% gels, treated with Amplify (Amersham), and fluorographed at -70 °C.

FIG. 3 *Cim*-correlated sequence polymorphism in *mtp2* from the RT1^a and RT1ⁱ (*cim*^a) and RT1^c and RT1^u (*cim*^b) MHC haplotypes. Rat sequences are also compared with the published human PSF2 sequence³⁴. Identity with *mtp2*^a is indicated by a minus sign. The sequence of PSF2 is identical to the human *RING11B* allele³⁵ except at position 379, where *RING11B* encodes a valine in place of isoleucine.

METHODS. Full-length coding sequences of the *mtp2* *c*, *i* and *u* alleles were obtained. Oligo(dT)-primed cDNA was amplified by PCR from poly(A) messenger RNA purified from conA-activated lymphocytes of the congenic and MHC recombinant rat strains PVG.R1, PVG-RT1ⁱ (LEW) and PVG.R8 respectively^{15,16,18}. cDNA was amplified between two overlapping pairs of primers based on the sequence of *mtp2*^a to obtain overlapping 5' and 3' halves of each sequence. All sequences were obtained in both orientations by means of consecutive sequencing primers based on the *mtp2*^a sequence²³. Ambiguities and compressions were resolved using dITP when required. X marks an as-yet unresolved ambiguity in the *mtp2*ⁱ sequence at position 145. The *mtp2*^a sequence is deposited in the EMBL Data Library, accession number X63854 (rat *mtp2* cDNA).

U	---	Y	---	R	---		60
C	---	Y	---	R	---	V	60
L	---						60
A	M	A	L	S	H	P	60
PSF2	-	R	-	P	-	D	60
U	---						120
C	---						120
L	---						120
A	L	V	G	T	F	L	120
PSF2	F	-	L	-	L	-	120
U	---						180
C	---						180
L	---						180
A	P	A	G	A	Q	E	180
PSF2	-	P	-	Q	-	D	180
U	---						240
C	---						240
L	---						240
A	G	G	D	F	D	P	240
PSF2	-	H	-	F	-	C	240
U	---						300
C	---						300
L	---						300
A	F	Q	E	T	K	T	300
PSF2	-	T	-	N	-	P	300
U	---						360
C	---						360
L	---						360
A	P	L	T	A	A	E	360
PSF2	-	F	-	T	-	E	360
U	---						420
C	---						420
L	---						420
A	R	C	R	Q	L	W	420
PSF2	Q	-	Y	-	R	-	420
U	---						480
C	---						480
L	---						480
A	H	H	V	Q	N	L	480
PSF2	S	Y	-	T	-	I	480
U	---						540
C	---						540
L	---						540
A	P	E	K	P	V	L	540
PSF2	-	D	-	R	-	E	540
U	---						600
C	---						600
L	---						600
A	Y	L	H	R	Q	V	600
PSF2	-	S	-	S	-	R	600
U	---						660
C	---						660
L	---						660
A	G	E	K	S	Q	L	660
PSF2	-	A	-	D	-	V	660
U	---						703
C	---						703
L	---						703
A	H	R	L	B	T	V	703
PSF2	-	Q	A	-	R	-	703

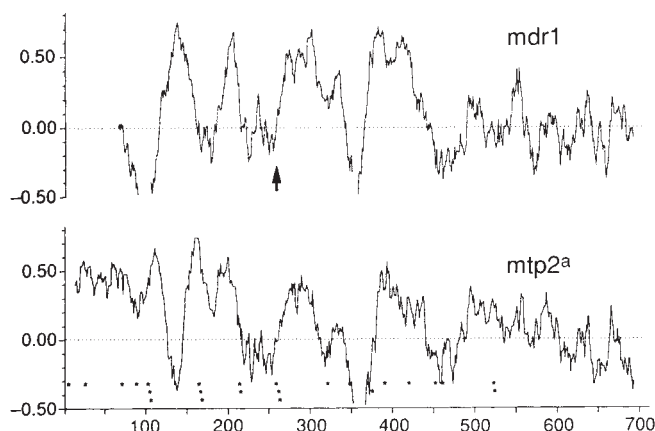


FIG. 4 Hydrophobicity plot of *mtp2^a*, with the *cim*-associated coding substitutions indicated by asterisks, aligned with the hydrophobicity plot of the first homology unit of human multidrug resistance (*mdr1*)⁴⁶ showing the position of Val 185, in which mutation to Gly 185 is correlated with resistance to colchicine³⁹ (arrow).

METHODS. Hydrophobicity plots were generated using the programme AMPHI as described¹⁹. The *mdr1* and *mtp2^a* sequences are aligned on the glycine of the Gly-Lys-Ser sequence of the Walker A nucleotide-binding motif.

consider it unlikely that these peaks all represent peptides derived from the sequence of the transfected *mtp2^a* itself because of the similarity to the A⁺ con A blast data already described, the persistence of the A⁺/A⁻ difference during presentation of the male H-Y antigen¹⁵, in which *mtp2^a*-derived peptides should play no part, and evidence that *mtp2^a* restores the presentation of viral antigen by the mutant cell line RMA-S (ref. 23). Finally, *cim*-correlated sequence substitutions make *cim^a* more hydrophobic than *cim^b*, yet the peptide profiles associated with *mtp2^a* transfectants are more hydrophilic.

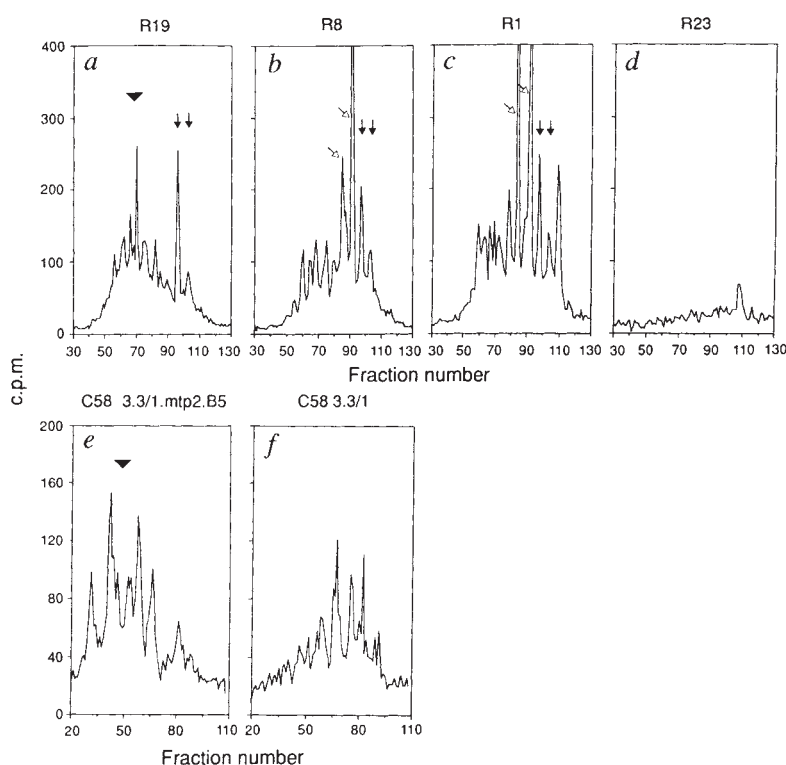
Attempts to demonstrate *cim*-dependent effects on the antigenicity or intracellular assembly of other RT1.A class I alleles, including RT1.A^c and RT1.A^u, have so far given negative results (ref. 18; unpublished observations). The implication is therefore that *mtp2* polymorphism has a selective effect on peptide loading of the RT1.A^a allelic product. We have performed preliminary experiments comparing peptides internally labelled with ³⁵S-Met from the endogenous RT1.A^u molecules expressed by the

C58-3.3/1.*mtp2^a* and C58-3.3/1 cells. The RT1.A^u-associated peptide profiles obtained so far from the two cell lines appear virtually identical to each other (data not shown).

Discussion

The proposed role for the MHC-linked ABC transporters in peptide transport from the cytosol to the endoplasmic reticulum^{19-22,28,36,37} remains to be demonstrated formally. Recent evidence opposed to this view indicates that short peptides can enter microsomal vesicles *in vitro* in the absence of ATP³⁸. Furthermore, entry of peptides is unimpaired in microsomes obtained from the T2 human cell line, which has a deletion spanning both MHC-linked transporter genes. Nevertheless, the functional arguments in favour of a role in peptide transport are strong. First, functional loss of either transporter chain by mutation causes an arrest in class I assembly and antigen presentation^{8,23-27}. Second, although such mutant cells are virtually unable to load peptides generated in the cytosol either by

FIG. 5 Reverse-phase HPLC analysis of radiolabelled peptides eluted from immunoprecipitated RT1.A^a class I molecules. *a*, RT1.A^a-associated peptides from PVG.R19 (*cim^a*, *mtp2^a*, A⁺) con A blasts. *b*, RT1.A^a-associated peptides from PVG.R8 (*cim^b*, *mtp2^a*, A⁺) con A blasts. *c*, RT1.A^a-associated peptides from PVG.R1 (*cim^b*, *mtp2^c*, A⁺) con A blasts. *d*, Control signal from non-RT1.A^a-expressing PVG.R23 (RT1.A^u) con A blasts. Peptides eluted from the *cim^a*, A⁺ strain cells (*a*) characteristically elute earlier in the acetonitrile gradient than those from *cim^b*, A⁺-strain cells (*b* and *c*). Peaks annotated with arrows are described in the text. *e*, RT1.A^a-associated peptides from C58-3.3/1.*mtp2^a*.B5 cells. *f*, RT1.A^a-associated peptides from C58-3.3/1 cells. RT1.A^a-associated peptides again elute earlier in the acetonitrile gradient in the presence of *mtp2^a* (*e* and *f*). METHODS. Labelling and isolation of class I-associated peptides was essentially according to refs 2 and 30. Briefly, 50 × 10⁶ con A blasts were labelled for 5 h with 37 MBq each of ³H-Leu and ³H-Tyr (*a-d*), or transfectant cells for 12 h with 18.5 MBq of ³⁵S-methionine (*e* and *f*). RT1.A^a was immunoprecipitated using MAC30 and protein A-Sepharose (Pharmacia). Isolated complexes were boiled in 10% acetic acid and the peptide fraction isolated by centrifugation through a Centricon 10 unit (Amicon). HPLC analysis was done on a Waters 600 unit with a YMC C₁₈ column at a flow rate of 2 ml min⁻¹ (*a-d*), or on a Pharmacia Smart system (*e* and *f*) using a micro-reverse-phase C₂/C₁₈ column at a flow rate of 200 μl min⁻¹. Columns were run in 0.1% trifluoroacetic acid in H₂O and eluted in 0.1% trifluoroacetic acid in acetonitrile. The amount of radioactivity in each fraction was counted; data represent fractions collected over a linear acetonitrile gradient of 10–50%. The abscissa indicates fraction number during column elution, the ordinate indicates counts per minute (c.p.m.) detected for ³H-labelled samples (*a-d*) and ³⁵S-labelled samples (*e* and *f*).



infection or transfection, the defect can be bypassed if a known peptide epitope is added at the cell surface⁸, or is introduced by transfection with an added leader sequence that can target it independently to the endoplasmic reticulum by the signal-dependent translocation pathway³⁹. The results we present here indicate a specific relationship between the transporter and the resulting peptides found within the class I binding groove, a relationship which is highly suggestive of a direct interaction between the transporter and the peptides. But this result still fails to distinguish between peptide transport across the endoplasmic reticulum membrane and chaperonin-like peptide binding during the delivery process as the key function of the MHC-linked transporters.

The RT1.A^a class I molecule appears to be peculiarly sensitive to the difference between the alleles of the *mtp2* transporter chain. It would appear that polymorphic variation in *mtp2* is directed towards the supply of peptides to specific class I alleles that may have unusual requirements for assembly. What these requirements are for RT1.A^a, and exactly how the peptides assembled with the A^{a+} and A^{a-} forms of RT1.A^a differ, are presently under investigation. Furthermore, the polymorphic sequence variation of *mtp2* provides an opportunity to perform structure/function analysis of differential peptide loading. Fragment exchanges and site-directed mutagenesis between *mtp2^a* and *mtp2^u* are in progress to localize functionally significant residues. Few clues are available to predict the outcome of sequence exchanges. In human mdrl, however, a conservative glycine/valine alteration adjacent to a predicted membrane-spanning segment changes the preferred substrate specificity in favour of colchicine resistance (Fig. 4)⁴⁰. And in the human gene encoding the cystic fibrosis transmembrane conductance regulator, mutation of lysines, again in or near putative membrane-spanning sequences, alters anion selectivity⁴¹.

Our reconstruction of the *cim* effect by transfection of *mtp2^a* into a rat T-cell line failed to reproduce the dominance of the *cim^a* over the *cim^b* phenotype invariably seen in cells derived from *cim^a* × *cim^b* heterozygous animals; in the *mtp2^a* transfectants expression of the A^{a-} form was not extinguished (Fig. 1e and f). This result was quantitatively reproduced in both bulk transfectants and multiple independent clones, suggesting a stable property of the system. Thus we have to consider what the molecular basis of *cim^a* dominance might be. We previously

proposed that the slow intracellular transit of RT1.A^a in the presence of *cim^b* alone could result from delivery of peptides ill-suited for assembly with this particular class I allelic product¹⁸, because assembly and subsequent transit depend on successful peptide loading. In a heterozygous cell a generous supply of appropriate peptides delivered by the *cim^a* transporter would pre-empt assembly with peptides delivered by the *cim^b* transporter, resulting in effective dominance of *cim^a* over *cim^b*. The functional transporter presumably contains both *mtp1* and *mtp2* chains, and polymorphism in the *mtp1* chain might also contribute to the *cim* effect. But in the single-chain transfectants reported here, the *mtp2^a* chain can presumably form associations only with the endogenous *mtp1^u* chain. These heterologous transporter complexes may be less efficient than the homologous *mtp1^a/mtp2^a* complex in delivering RT1.A^a compatible peptides, and therefore unable to diminish A^{a-} expression below the level of detection by CTL. We are therefore performing double transfections of both *mtp1^a* and *mtp2^a* to explore this question directly, as well as looking for *cim*-correlated *mtp1* polymorphism.

How special is the *cim* phenomenon? So far no functional polymorphism on the scale of *mtp2* has been documented in the homologous genes of human or mouse^{34,35}. The mouse Qa-2 non-classical class I product has CTL determinants dependent on a gene mapping in the region of the transporters⁴², but little else is known about this phenomenon. If polymorphism does exist in the human transporter genes then its relevance to autoimmune diseases will be a key issue. It is possible that relatively weak associations of certain diseases with MHC class II alleles might be strengthened if in addition polymorphism in the transporter region is considered. But the strongest suggestion of *cim*-like phenomena in man has recently been reported in the spondyloarthropathy-associated class I allele, *HLA-B27*. There is evidence suggesting that *HLA-B27* may be subject to biosynthetic and antigenic modification by another MHC-linked locus⁴³. Furthermore, these effects are associated with striking changes in the typical hydrophobicity of peptides loaded into B27 molecules that closely resemble those described in this study (L. Pazmany *et al.*, personal communication). It will be of interest to examine the sequences of the MHC-linked transporters and the peptide contents of the class I molecules in these systems. □

Received 2 March; accepted 27 March 1992.

1. Bjorkman, P. J. *et al.* *Nature* **329**, 507–512 (1987).
2. Van Bleek, G. M. & Nathenson, S. G. *Nature* **348**, 213–216 (1990).
3. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Nature* **353**, 321–325 (1991).
4. Jardetzky, T. S. *et al.* *Nature* **353**, 326–329 (1991).
5. Falk, K. *et al.* *Nature* **351**, 290–296 (1991).
6. Townsend, A. R. M. *et al.* *Cell* **39**, 13–25 (1984).
7. Townsend, A. & Bodmer, H. A. *Rev. Immun.* **7**, 601–624 (1989).
8. Townsend, A. *et al.* *Nature* **340**, 443–448 (1989).
9. Townsend, A. *et al.* *Cell* **62**, 285–295 (1990).
10. Elliott, T., Cerundolo, V., Elvin, J. & Townsend, A. *Nature* **351**, 402–409 (1991).
11. Silver, M. L., Parker, K. C. & Wiley, D. C. *Nature* **350**, 619–622 (1991).
12. Rötzschke, O. *et al.* *Nature* **348**, 252–254 (1990).
13. Howard, J. C., Butcher, G. W., Galfre, G. & Milstein, C. *Contemp. Topics Microbiol. Immun.* **81**, 54–60 (1978).
14. Livingstone, A. M., Butcher, G. W. & Howard, J. C. *Transplant. Proc.* **15**, 1557–1559 (1983).
15. Livingstone, A. M. *et al.* *J. exp. Med.* **170**, 777–795 (1989).
16. Livingstone, A. M. *et al.* *Immunogenetics* **34**, 157–163 (1991).
17. Powis, S. J., Howard, J. C. & Butcher, G. W. *Transplant. Proc.* **22**, 2517–2518 (1990).
18. Powis, S. J., Howard, J. C. & Butcher, G. W. *J. exp. Med.* **173**, 913–921 (1991).
19. Deverson, E. V. *et al.* *Nature* **348**, 738–741 (1990).
20. Monaco, J. J., Cho, S. & Attaya, M. *Science* **250**, 1723–1726 (1990).
21. Trowsdale, J. *et al.* *Nature* **348**, 741–744 (1990).
22. Spies, T. *et al.* *Nature* **348**, 744–747 (1990).
23. Powis, S. J. *et al.* *Nature* **354**, 528–531 (1991).
24. Spies, T. & DeMars, R. *Nature* **351**, 323–324 (1991).
25. Kelly, A. *et al.* *Nature* **355**, 641–644 (1992).

26. Spies, T. *et al.* *Nature* **355**, 644–646 (1992).
27. Attaya, M. *et al.* *Nature* **355**, 647–649 (1992).
28. Townsend, A. R. M., Gotch, F. M. & Davey, J. *Cell* **42**, 457–467 (1985).
29. Garrett, T. P. J. *et al.* *Nature* **342**, 692–696 (1989).
30. Van Bleek, G. M. & Nathenson, S. G. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11032–11036 (1991).
31. Falk, K., Rötzschke, O. & Rammensee, H.-G. *Nature* **348**, 248–251 (1990).
32. Silva, A. *et al.* *Immun. Rev.* **76**, 105–129 (1983).
33. Rada, C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 2167–2171 (1990).
34. Bahram, S. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 10094–10098 (1991).
35. Powis, S. H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **89**, 1463–1467 (1992).
36. Parham, P. *Nature* **348**, 674–675 (1990).
37. Robertson, M. *Nature* **353**, 300–301 (1991).
38. Lévy, F., Gabathuler, R., Larsson, R. & Kvist, S. *Cell* **67**, 265–274 (1991).
39. Anderson, K. *et al.* *J. exp. Med.* **174**, 489–492 (1991).
40. Choi, K., Chen, C., Kriegler, M. & Roninson, I. B. *Cell* **53**, 519–529 (1988).
41. Anderson, M. P. *et al.* *Science* **253**, 202–205 (1991).
42. Murphy, D. B. & Homer, R. J. in *H-2 Antigens Genes, Molecules, Function* (ed. David, C. S.) 359–369 (Plenum, New York, 1987).
43. Pazmany, L. *et al.* *J. exp. Med.* **175**, 361–369 (1992).
44. Gunning, P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 4831–4835 (1987).
45. Butcher, G. W. *J. Immunogenet.* **14**, 163–176 (1987).
46. Chen, C.-J. *et al.* *Cell* **47**, 381–389 (1986).

ACKNOWLEDGEMENTS. We thank A. McMichael and J. Trowsdale for communication of results before publication, G. van Bleek for advice on peptide elution, and A. Hutchings and E. Tredgett for assistance with cell culture. S.J.P. is supported by the Arthritis and Rheumatism Council.

The crystal structure of diphtheria toxin

Seunghyon Choe*, Melanie J. Bennett*, Gary Fujii, Paul M. G. Curmi, Katherine A. Kantardjieff, R. John Collier† & David Eisenberg*

* Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90024, USA

† Department of Microbiology and Molecular Genetics, Harvard Medical School, Shipley Institute of Medicine, Boston, Massachusetts 02115, USA

The crystal structure of the diphtheria toxin dimer at 2.5 Å resolution reveals a Y-shaped molecule of three domains. The catalytic domain, called fragment A, is of the $\alpha + \beta$ type. Fragment B actually consists of two domains. The transmembrane domain consists of nine α -helices, two pairs of which are unusually apolar and may participate in pH-triggered membrane insertion and translocation. The receptor-binding domain is a flattened β -barrel with a jelly-roll-like topology. Three distinct functions of the toxin, each carried out by a separate structural domain, can be useful in designing chimaeric proteins, such as immunotoxins, in which the receptor-binding domain is substituted with antibodies to target other cell types.

DIPHtheria was a major cause of death among children until mass immunization against diphtheria toxin (DT) reached the general public in the late 1920s. DT is secreted as a single polypeptide chain of 535 residues from *Corynebacterium diphtheriae* that has been lysogenized by a bacteriophage carrying the DT gene¹. Mild trypsinization and reduction of DT *in vitro* generates two fragments, fragment A (amino terminal, about 21K) and fragment B (carboxy terminal, about 37K), as a result of cleavage at residue 190, 192 or 193 (refs 2, 3). A similar proteolytic cleavage ('nicking') occurs *in vivo* before or soon after the toxin binds to a sensitive cell⁴. Fragment B of the toxin binds the protein to receptors on the cell surface and promotes transfer of fragment A to the cytoplasm. Fragment A in the cytoplasm catalyses the transfer of the ADP-ribosyl group of NAD⁺ to elongation factor 2 (EF-2). This inactivates EF-2, stopping protein synthesis and killing the target cell. Introduction of a single molecule of fragment A into the cytoplasm can kill a cell⁵. Studies are underway to create therapeutically active chimaeric proteins, with a part of DT linked to an antibody or other ligand that can recognize and deliver the toxin's lethal activity to a specific target cell⁶⁻⁹.

Beyond its interest as a toxic enzyme, DT has been studied to understand the mechanism by which it translocates its enzymatically active fragment A into cells. After binding to a receptor, the toxin is endocytosed and delivered to endosomes, where conditions are acidic. At a threshold pH of ~ 5.0 the toxin undergoes a conformational change, which promotes insertion and formation of an ion-selective channel in the membrane, and fragment A is translocated and released into the cytoplasm¹⁰⁻¹². The relationship between pore formation and transfer of fragment A across the endosomal membrane remains uncertain. Hydrophobic segments in the central part of fragment B have been implicated in the channel formation^{13,14}. The membrane translocation machinery of DT, therefore, can be considered as a simple translocation system consisting of a single protein, which may illuminate the protein secretory apparatus of cells.

Structure determination

The structure is based on analyses of form 1, form 3 and form 4 crystals. Form 1 crystals of DT complexed with adenylyl-3',5'-uridine monophosphate (ApUp) belong to triclinic space group *P*1 with unit cell dimensions of $a = 70.4$ Å, $b = 70.6$ Å, $c = 65.4$ Å, $\alpha = 94.9^\circ$, $\beta = 91.0^\circ$ and $\gamma = 99.6^\circ$, with two chains per

asymmetric unit. This dimeric asymmetric unit is consistent with the discovery, after our initial report of crystallization¹⁵, that the crystals were of a dimeric form of DT sometimes found in crude or purified preparations of the protein. Dimeric DT itself is not toxic, presumably because it does not bind to receptors, but it slowly dissociates to fully toxic monomers¹⁶. The dimer may represent a conformationally altered form of the biologically active monomeric toxin. Irreproducible crystallization conditions for obtaining form 1 crystals hampered crystallographic studies of structure determination until three new crystal forms were obtained¹⁷. Form 3 and form 4 belong to monoclinic space group *C*2, with unit cell dimensions for form 3 of $a = 107.3$ Å, $b = 91.7$ Å, $c = 66.3$ Å and $\beta = 94.7^\circ$, and for form 4 of $a = 108.3$ Å, $b = 92.3$ Å, $c = 66.1$ Å and $\beta = 90.4^\circ$. In both of these forms there is one DT chain per asymmetric unit and pairs of DT chains are related by a 2-fold rotation axis.

The initial model was based on the structure determination of form 4 crystals at 3.0 Å resolution, using the multiple isomorphous replacement (MIR) method followed by solvent flattening¹⁸. With the initial model, the structures of form 1 and form 3 were readily solved by molecular replacement^{19,20}. Single isomorphous replacement (SIR) phases were also obtained for form 3. Native data were then collected to 2.5 Å resolution, and the model was rebuilt into 2.5 Å maps with form 3 (SIR) and form 4 (MIR) after the phases had been extended and modified by the method of Zhang and Main²¹. This was followed by real-space density averaging between two forms. Sequence fitting was difficult in the ~ 120 C-terminal residues (part of the receptor-binding or R domain) where the most ambiguous regions were near residues 408 and 510. Some of the useful markers in the density maps were Trp 50, Trp 153, Trp 281, Trp 398, a 5-residue segment of Met 178, Tyr 179, Glu 180, Tyr 181, Met 182, a 4-residue cluster of Phe 355, Tyr 358, His 372, Tyr 375, a cluster of Tyr 514, Phe 530, Phe 531 with big side chains near the C terminus (Fig. 1c), and two disulphide bonds between Cys 186 and Cys 201 and Cys 461 and Cys 471. An initial improper fitting in the R domain was detected by profile window plots²² and then corrected. Iterative cycles of refinement were done independently at 2.5 Å for each dataset. The atomic model for each form is essentially identical except for crystal packing. Details of phase modification and refinement will be described elsewhere. Assessment of the accuracy of the model rests on the fit of the model to the MIR and density-modified maps, crystallographic *R* factors, real-space *R* factors²³, the free *R* value²⁴, which is only 4% higher than the crystallographic *R* factor, and profile window plots²². At the present stage of refinement, the agreement of the atomic models to crystallographic data is characterized by *R* factors of 21.1, 21.6 and

Present address: Vestar, Inc., San Dimas, California 91773, USA (G.F.); School of Physics, University of New South Wales, Kensington, Australia (P.M.G.C.); Department of Chemistry and Biochemistry, California State University, Fullerton, California 92634, USA (K.A.K.).

TABLE 1 Statistics for X-ray data collection, phase determination and refinement

Native data		Overall (R_{scale}^*)	10–4.0	4.0–3.5	3.5–3.0	3.0–2.5 Å
Form 4	Total	36,758 (11.9)	26,897	4,977	4,884	
	Unique (% complete)	10,875 (83)	5,190 (99)	2,414 (92)	3,271 (63)	
Form 4 (new)	Total	35,897 (6.1)				
	Unique (% complete)	18,665 (84)	5,195 (99)	2,673 (98)	4,268 (82)	6,529 (72)
Form 3	Total	61,009 (7.6)	21,984	15,368	10,573	13,084
	Unique (% complete)	19,912 (90)	5,231 (100)	2,682 (98)	4,603 (88)	7,396 (82)
Form 1	Total	66,464 (7.5)	22,245	21,118	15,277	7,824
	Unique (% complete)	25,854 (68)	6,523 (96)	7,665 (92)	7,102 (76)	4,574 (35)
Derivatives		Overall	10–4.6	4.6–3.6	3.6–3.0	3.0–2.8
Form 4 KOS	Unique (R_{scale}^*)	11,765 (9.16)				
	$R_c^{\dagger}(\text{fh}/e)$	0.66 (1.23)	0.66 (1.29)	0.62 (1.18)	0.75 (1.28)	0.80 (1.06)
CNP	Unique (R_{scale}^*)	12,255 (12.0)				
	$R_c^{\dagger}(\text{fh}/e)$	0.70 (1.06)	0.68 (1.33)	0.72 (0.93)	0.72 (0.97)	0.66 (1.11)
KNP	Unique (R_{scale}^*)	8,164 (8.32)				
	$R_c^{\dagger}(\text{fh}/e)$	0.71 (1.00)	0.72 (0.87)	0.67 (1.18)	0.75 (1.33)	
CAP	Unique (R_{scale}^*)	7,552 (15.3)				
	$R_c^{\dagger}(\text{fh}/e)$	0.71 (1.28)	0.70 (1.54)	0.72 (1.12)	0.88 (1.20)	
KAP	Unique (R_{scale}^*)	10,152 (12.26)				
	$R_c^{\dagger}(\text{fh}/e)$	0.81 (1.26)	0.81 (1.43)	0.71 (1.19)	0.75 (0.89)	
GCL	Unique (R_{scale}^*)	6,595 (11.90)				
	$R_c^{\dagger}(\text{fh}/e)$	0.70 (1.10)	0.69 (1.13)	0.66 (1.09)	0.50 (1.05)	
Form 3 KOS	Unique (R_{scale}^*)	11,435 (13.57)				
	$R_c^{\dagger}(\text{fh}/e)$	0.54 (1.10)	0.56 (1.32)	0.60 (0.74)		
Refinement		Form 1	Form 3	Form 4		
R factor§ (6–2.5 Å)		0.211	0.216	0.219		
r.m.s. bond (Å)		0.021	0.021	0.021		
r.m.s. angle		4.54	4.40	4.48		
r.m.s. dihedral		26.4	25.9	26.1		

Crystal forms 1, 3 and 4 were used for the current study¹⁷. Diffraction data were collected on a Rigaku AFC-6 diffractometer operating at 8.5 kW, equipped with a two-panel area detector of Xuong-Hamlin design (San Diego Multiwire Systems). Images were recorded as 0.1° oscillation frames, integrated and merged into batches of 50 frames (5°). Integrated intensities were scaled and merged by the FOURIER scaling method⁴². Form 4 native and derivative data were later collected to 2.5 Å with a RAXIS imaging plate system. Heavy atom derivatives: KOS, K_2OsO_4 , soaked for 3 days at saturated concentration in artificial mother liquor (12% PEG 8000, 0.43 M NaCl, 43 mM Tris-HCl, pH 7.8); CNP, 4-chloro-2-nitro-mercury phenol, soaked for 5 days at saturated concentration in artificial mother liquor; KNP, 1 to 1 mixture of KOS and CNP; CAP, trans-dichlorodiamine Platinum (II), soaked for 3 days at 2 mg ml^{−1} in artificial mother liquor; KAP, 1 to 1 mixture of KOS and CAP; GCL, HgCl_2 , soaked for 3 days at 2 mg ml^{−1} in artificial mother liquor. Heavy atom parameters were refined and MIR phases calculated using the program HEAVY⁴³. We initially obtained the Os derivative for form 3 crystals. From electron density maps based on the SIR phases after solvent flattening at 3.5 Å resolution, the shape of the molecule was interpreted to have three domains. But secondary structures were not easily interpretable and the course of the polypeptide chain was difficult to determine. A search for additional heavy atom derivatives was hampered by the lack of good quality crystals of form 3. We therefore shifted our efforts to form 4 crystals. MIR phases for form 4 were obtained from six heavy atom derivatives using isomorphous differences and anomalous differences. The Os and Pt derivatives were solved by isomorphous difference Patterson functions, and the Hg derivative by a difference Fourier synthesis. Os derivatives of form 4 and form 3 have the same single site binding. Solvent flattening: initial electron density maps of form 4 were calculated at 3.0 Å resolution, with phases modified using an iterative solvent flattening procedure¹⁸ including phases extended to 3.0 Å from 3.2 Å by the Wang phase extension algorithm¹⁸. A solvent volume of 45% was used to ensure that all protein density was included in the protein mask, somewhat smaller than the 57% estimated from the M_r . From these maps, all secondary structures were identified and an initial model was built using a polyaniline chain. Model building was expedited with the program FRODO⁴⁴ and the fragment-fitting routines of the program O²³. Starting with α -carbon coordinates that were manually built, main chain atoms were added using a database of 34 well refined protein structures. Then side chains were added using a rotamer database⁴⁵. Refinement: this initial model was adjusted by visual inspection of density maps before it was refined by the simulated annealing protocol of the program XPLOR⁴⁶. The relative orientations of DT in forms 1, 3 and 4 were determined by a Patterson-space rotation and translation search of the refined form 4 model against form 1 and form 3 data. Two top solutions (9 σ) for form 1 data correspond to two DT chains related by noncrystallographic symmetry in the asymmetric unit. The transformation from form 4 to form 1 is essentially a change of coordinate system from $C2$ to $P1$, where the crystallographic rotation axis of $C2$ becomes a noncrystallographic rotation symmetry axis of $P1$ that is nearly parallel to the (110) axis of $P1$. One top solution (7 σ) for form 3 corresponds to a rotation of less than 0.5° in all three directions. The transformation from form 4 to form 3 is essentially a 5 Å translation along the a axis. This result is consistent with the observation that the average absolute difference of the amplitudes of structure factors of $0kl$ reflections between form 3 and form 4 is 15%, whereas those between $h0l$ or $hk0$ reflections between form 3 and form 4 are almost random ($R=48\%$). Also, when the model was superimposed on the solvent-flattened electron density maps of form 3 based on the SIR phases, most of the secondary structures were recognized with the model as a guide. Real-space averaging of densities between form 4 and form 3 with MIR and SIR phases at 3.0 Å improved the density maps at this stage. Subsequently, experimental phases were extended to 2.5 Å by the algorithm based on solvent flattening, histogram matching, and Sayre's equation²¹ for form 3 and form 4. Form 3 maps at 2.5 Å were again skewed and averaged with form 4 maps. These were the most interpretable maps. Refinement of the atomic model was done independently for form 1, form 3 and form 4 with all observed data having F_{obs} greater than $1\sigma(F_{\text{obs}})$ between 6 and 2.5 Å.

* $R_{\text{scale}} = \sum (|I_i - I_j|) / \sum (I_{\text{av}})$ where I_i and I_j are the i th and j th measurements of the equivalent reflections⁴².

† R_c is the Cullis R factor for centric reflections.

‡ fh/e is the phasing power, fh , the mean amplitude of heavy atom structure factors divided by e , the r.m.s. lack-of-closure error.

§ R factor = $\sum (|F_{\text{obs}} - F_c|) / \sum (F_{\text{obs}})$ where F_{obs} and F_c are the structure factors observed and calculated from the model, respectively. The R factors for all forms increased by about 1.9% when a single temperature factor was used for all atoms.

21.9%, respectively, for form 1, form 3 and form 4 for all observed data having F_{obs} greater than $1\sigma(F_{\text{obs}})$ between 6 and 2.5 Å resolution.

The final model consists of 4,137 non-hydrogen atoms with individual isotropic temperature factors. The model also includes ApUp in the active site cleft of the catalytic (C) domain, but no solvent atoms. There are poorly defined regions in the electron density maps where main chain densities for residues 170–172, 190–195, 389–390, and 500–503 are not well defined. Residues 190–195 are part of the protease-sensitive region of the first disulphide loop, where nicking occurs; this region may be intrinsically flexible as may be the loop between the transmembrane (T) and R domains, which includes residues 389–390. Table 1 summarizes aspects of data collection, phase determination and refinement.

Structure of DT

DT consists of three abutting domains that are connected by interdomain linkers. The N-terminal C domain, middle T domain, and C-terminal R domain consist of residues 1–193, 205–378 and 386–535, respectively. Schematically, DT is Y-shaped with the base formed by the T domain, one arm of the Y formed by the C domain, and the other arm formed by the R domain. The Y is about 90 Å high, 50 Å across the top of the Y, but only 30 Å thick (Fig. 1).

Each of the three domains has a distinctive fold. The C domain is a mixed structure of eight β -strands (CB1–CB8) and seven α -helices (CH1–CH7). The eight β -strands form two β -sheets of three and five strands each. These β -sheets form a core that is surrounded by seven short helices. The overall folding of the C domain is similar to that of *Pseudomonas aeruginosa* exotoxin A (ETA), especially near the active site²⁵, a result that had been foreshadowed by a weak similarity in amino-acid sequences^{26,27}. The folding of the active site region of *Escherichia coli* heat-labile enterotoxin also closely resembles that of ETA²⁸. The T domain contains nine helices (TH1–TH9) that are folded into three helix layers, each of which is formed by two or more antiparallel helices. A similar feature was observed in the structure of the channel-forming domain of colicin A²⁹. The R domain contains 10 β -strands (RB1–RB10), nine of which (RB2–RB10) build two β -sheets. These two β -sheets form a β -sandwich with a topology similar to a jelly-roll fold³⁰. The three-domain organization of DT is shared by two other bacterial toxins: ETA and δ -endotoxin from *Bacillus thuringiensis*³¹. The catalytic domains of DT and ETA are the closest among all these domains in their structures and functions.

Catalytic domain. We view the C domain as being formed from two β -sheet subdomains, which subtend the active site cleft (Fig. 2). These β -sheets are oriented roughly perpendicular to each other and form the core of the domain. One subdomain consists of β -strands CB2, CB4 and CB8, surrounded by α -helices CH2, CH3, CH6 and CH7. The other subdomain consists of β -strands CB1, CB3, CB5, CB6 and CB7 surrounded by helices CH1, CH4 and CH5. The two subdomains are connected by extended loops CL1–CL4, which link the two subdomains. These four loops seem to endow the potential for flexibility or even extension to a longer and narrower shape. Conceivably the C domain can assume this partially unfolded structure during membrane translocation.

The active site cleft of the C domain, identified by the binding of the dinucleotide ApUp, is formed primarily by β -strands CB2, CB3 and CB7, the short helix CH3, and the loop CL2, and is also bounded by β -strand RB6 of the R domain. Located within the active site cleft are the following residues: Glu 148 which is believed to play a key role in catalysis³², His 21³³ and Tyr 65³⁴, both of which have been implicated in NAD⁺ binding, and various other residues suggested to be at or near the active site (Gly 52^{32,35}, Trp 50³⁶, Lys 39³⁷, Lys 474³⁸). Least-squares superposition of the α -carbon coordinates of the C domains of DT and ETA yields an r.m.s. difference of 1.44 Å between 85

residues (16–33, 34–38, 49–66, 75–90, 91–96, 131–136, 147–164 of DT and 437–452, 454–458, 465–482, 493–508, 511–516, 540–545, 552–569 of ETA).

The rough position of the substrate NAD⁺ in the active site can be inferred, because the dinucleotide, ApUp, binds competitively with NAD⁺. The high affinity of ApUp (~ 0.3 nM as compared with ~ 8 – 16 μ M for NAD⁺¹⁶) may be a consequence of multiple contacts with the C domain and of salt bridges between the 3'-terminal phosphate of ApUp and the side chains of Thr 42 and Arg 458, the latter of which is a residue of the R domain. Although the structure of bound ApUp resembles that of NAD⁺, there are enough differences between the covalent structures of NAD⁺ and ApUp to make difficult the prediction of the conformation of NAD⁺ in the cleft. But assuming that the adenine phosphate portion of NAD⁺ binds in the same conformation as that of ApUp, we find that the nicotinamide ring will be positioned close to the site of the uridine ring. This

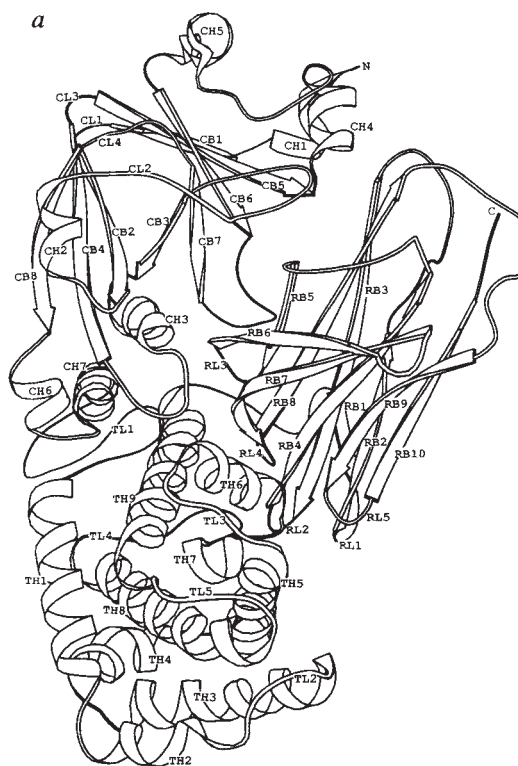
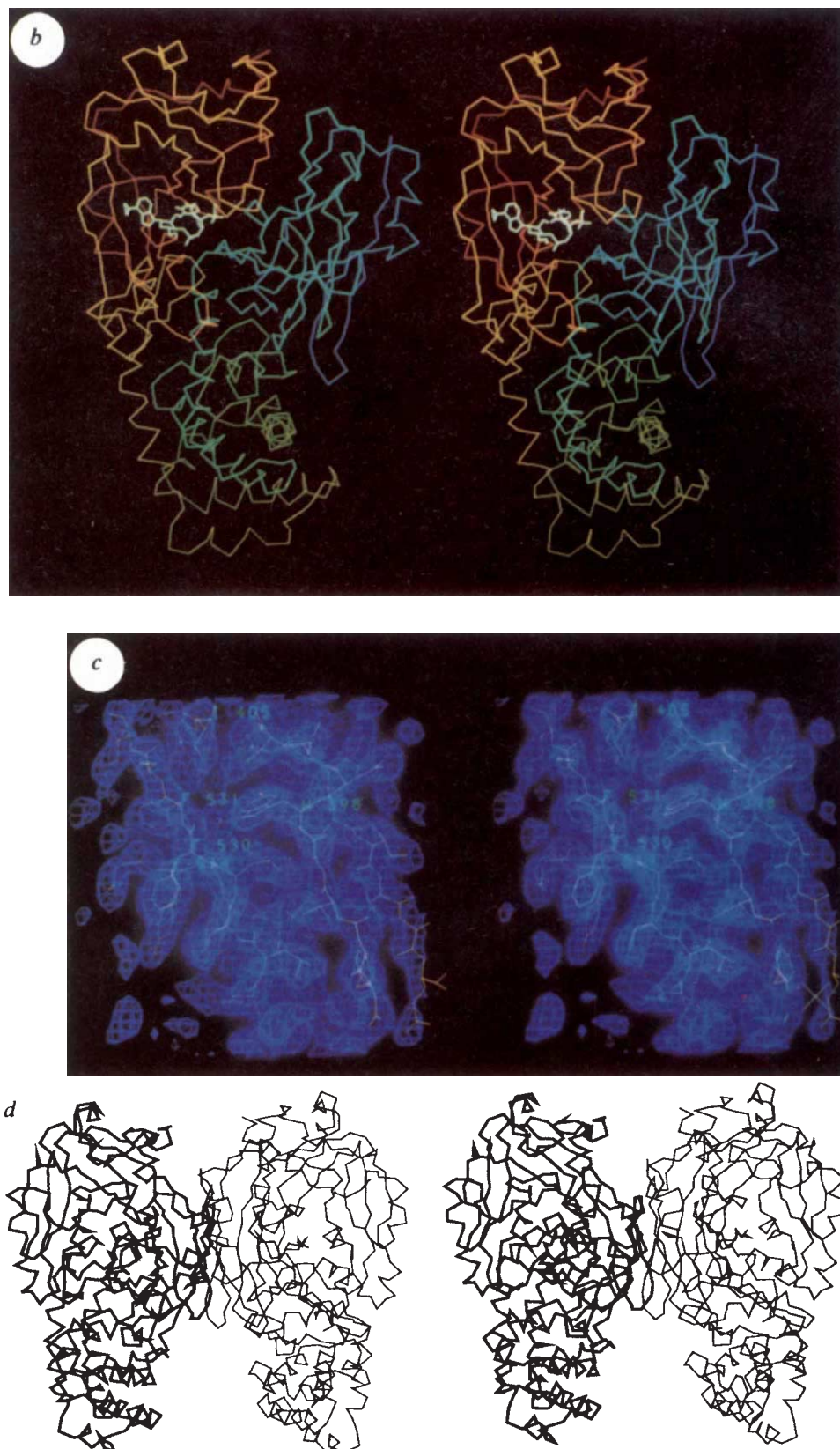


FIG. 1 a, Ribbon drawing of DT, labelling each secondary structural segment. The first letter denotes the domain: C for catalytic, T for transmembrane, and R for receptor-binding domains. The second letter denotes the secondary structure class: H for helix, B for β -strand, L for loop. The third symbol is the sequential number of each secondary segment from the N terminus of each domain. The ribbon program was kindly provided by J. Priestle. The residue numbers in each segment are as follows: CH1: 2–7, CB1: 11–14, CB2: 16–24, CH2: 28–34, CB3: 52–57, CH3: 58–66, CB4: 76–86, CB5: 88–96, CH4: 99–106, CH5: 120–126, CB6: 130–136, CB7: 147–152, CB8: 159–166, CH6: 168–173, CH7: 176–186; TH1: 205–221, TH2: 225–231, TH3: 238–257, TH4: 258–269, TH5: 274–288, TH6: 297–307, TH7: 310–315, TH8: 326–346, TH9: 356–378; RB1: 386–390, RB2: 393–399, RB3: 412–424, RB4: 428–438, RB5: 447–453, RB6: 455–465, RB7: 467–480, RB8: 483–495, RB9: 513–520, and RB10: 525–534. b, Stereo diagram of the α skeleton of DT from the same viewpoint as that of a. The colour changes gradually from red (N terminus) to blue (C terminus). An ApUp molecule (white) occupies the active site of DT. c, Stereo pairs of electron density maps calculated at 2.5 Å from ($2F_{\text{obs}} - F_c$) and the refined model phases. Maps are superimposed on the corresponding region of the refined model. d, The DT dimer observed in the form 4 crystal. The two monomers are related by a crystallographic 2-fold rotation axis, which is vertical. The molecule at the left (in thick line) has the same orientation as that in a.

places the nicotinamide ring adjacent to side chains of His 21, Tyr 65 and Glu 148.

Domain junctions. One of the two intramolecular disulphide bonds of DT bridges a handle-like loop TL1 on the molecular surface (Fig. 1*a*). This 14-residue loop (187–200) connects fragment A to fragment B; it is rich in Arg and known to be easily

nicked by proteases^{2,3}. Once this loop is nicked, fragment A and fragment B are covalently linked only by the disulphide bond. There is evidence that nicking plays a part in the cytotoxic action of DT⁴, and it is generally believed that nicked DT separates into free fragment A and fragment B when this disulphide bond is exposed to the reducing environment of the cytoplasm during



membrane translocation of the toxin. The second disulphide bond makes a nine-residue loop between residues 461 and 471 within fragment B. Residues near this loop (456, 458, 460, 472, 474) are also rich in positive charges and face the active-site cleft, probably forming the so-called phosphate-binding 'P-site' (ref. 39).

The structure suggests why whole DT is inactive in catalysing the ADP-ribosylation of EF-2 until the C domain dissociates, in the form of fragment A, from fragment B. The active site is formed at the interface between the C domain and the R domain (Fig. 1b). Entry to the active site is shielded by the 18-residue loop CL2 and the R domain. Thus, in whole DT, the approach of EF-2 ($M_r \sim 100K$) to the active site is blocked. The active site of whole DT remains accessible to NAD^+ , however, and catalyses the NAD-glycohydrolysis (a slow side reaction that is probably physiologically insignificant). The lack of secondary structural elements within loop CL2 may allow a substantial movement of main chain atoms of the loop, permitting substrate entry to the active site.

Transmembrane domain. A central unanswered question about DT is how the low-pH environment of the endosome triggers DT insertion into the endosomal membrane and how this insertion aids the translocation of the C domain into the cytoplasm. The structure of the T domain exhibits two features that suggest how it might experience pH-triggered insertion into the membrane. The first is that the T domain is entirely α -helical, similar to the known and proposed transmembrane proteins, and that some of the helices have hydrophobic characteristics more typical of transmembrane helices than of globular proteins⁴⁰. The nine helices are arranged more or less in three layers, each layer consisting of an antiparallel pair of helices. The two long, C-terminal helices, TH8 and TH9, are unusually apolar and constitute the central core layer. One flanking layer, made up of helices TH5-TH7, also contains hydrophobic helices, TH6 and TH7. The other layer, made up of helices TH1-TH3, is, in contrast, very hydrophilic even compared with globular proteins. The second noteworthy feature of the T domain is the acidic composition of the loops that connect pairs of these helices: both loop TL3 between helices TH5 and TH6, and loop TL5 between hydrophobic helices TH8 and TH9 contain a total of six Asp and Glu residues (Fig. 3a). At neutral pH, these loops are highly charged and water-soluble. But at acidic pH, these residues would be at least partially protonated, and hence more nearly neutral and membrane-soluble, especially near the surface of the membrane that has even higher concentration of protons due to the surface potential⁴¹. Thus the lower pH inside the endosome would tend to render these tip-shaped loops into

membrane-soluble 'daggers' that would lead the two apolar helix pairs into the membrane.

Other structural characteristics of the T domain suggest that it has the capacity to insert into the membrane and can assist the translocation of the C domain. The first is that the nearly parallel packing of the three helix layers would permit spreading on the membrane surface of the first helix layer (TH1-TH3) if other layers were inserted. This insertion would require local conformational changes in loops, but no alteration of the helices themselves. Also the pronounced hydrophobic asymmetry is compatible with the proposed rearrangement: 15 of 16 Lys and Arg residues and all six His residues of the T domain are located on the opposite side from the dagger tips (Fig. 3b), making the whole domain a hydrophobic dipole once the Asp and Glu residues are neutralized. In short, we propose that the hairpin loop TL5 and probably TL3 cross the membrane, where the Asp and Glu residues will once again be charged in the neutral pH of the cytoplasm.

Receptor-binding domain. The R domain is formed from two β -sheets. β -Strands RB2, RB3, RB5 and RB8 form a four-stranded β -sheet that faces a five-stranded β -sheet containing β -strands RB4, RB6, RB7, RB9 and RB10. RB6 interacts with both β -sheets through hydrogen bonds. The connection of the strands is such that the R domain is similar to the jelly-roll topology found in many proteins that are exclusively formed from antiparallel β -strands³⁰. Jelly-roll domains include viral coat proteins, tumour necrosis factor, and the receptor-binding domain of ETA. The R domain differs somewhat from a strict jelly-roll topology (Fig. 4) in having strand 2 in the 'front' sheet, and having strand 10 in the 'back'. The R domain also is reminiscent of an immunoglobulin variable domain, but differs from the immunoglobulin fold in having an 'insert' of strands 5 and 6 between 4 and 7, and also in lacking two short strands (C' and C'' in Fig. 4) between 4 and 5. The portion of the R domain that resembles a strict jelly-roll in topology is the right side as viewed in Fig. 4; and the portion that resembles the immunoglobulin variable domain is the left side, the side that is away from the rest of the DT monomer. Conceivably it is this immunoglobulin variable-like moiety that is involved in receptor recognition.

The DT dimer. Two monomers associate tightly to form a dimer with an interface between RB1/RB2 of one DT molecule and RB2/RB1 of the other DT molecule related by 2-fold rotational symmetry (Fig. 1d). This interface is one of three major protein-protein contacts in crystal packing and involves three hydrogen bonds per monomer. These hydrogen bonds are well defined because they are formed between main chain N and C atoms

FIG. 2 Stereo pair of the $C\alpha$ skeleton of the C domain. The entrance to the active site is at the lower right. The four loops, CL1 to CL4, are highlighted. Notice that they form a hinge which may permit the C domain to form a more elongated structure.



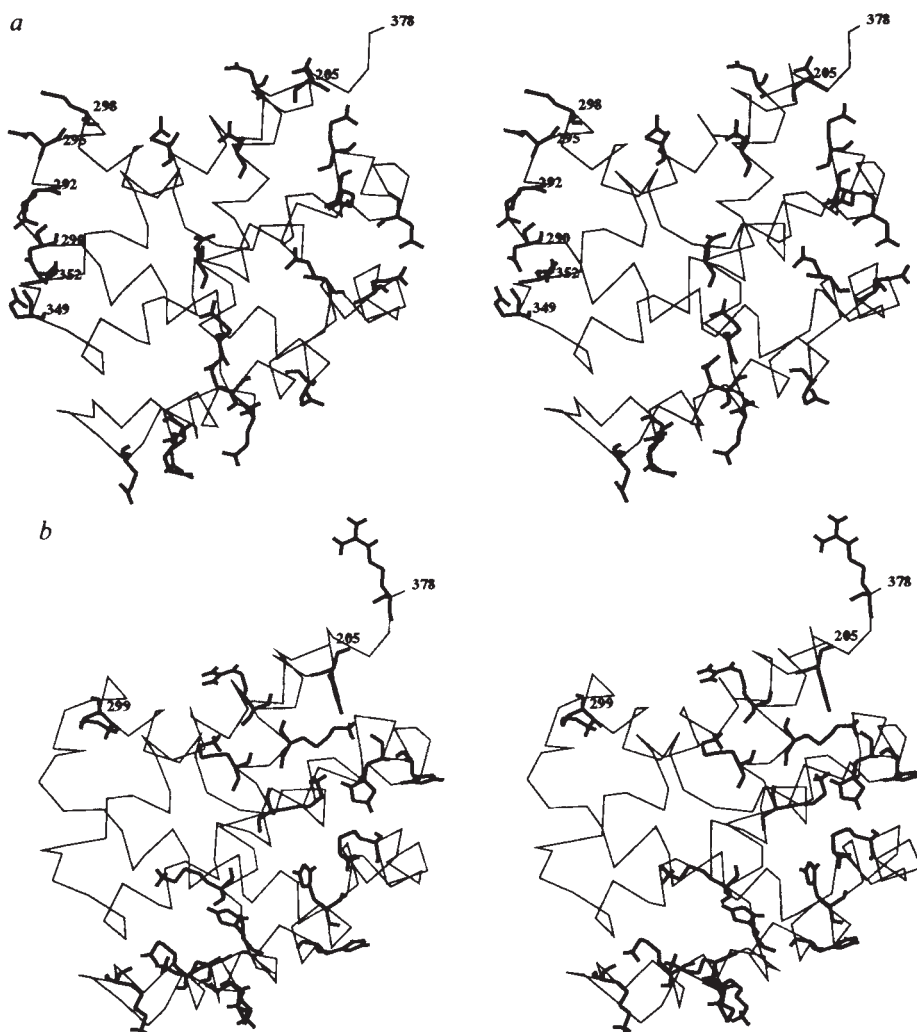


FIG. 3 Stereo pair of the $C\alpha$ skeleton of the T domain, with the direction of view from the right side of DT in Fig. 1. Helix TH1 lies at the back, starting at residue 205. Helix TH2 runs to the left at the bottom, followed by a turn and helix TH3 running to the right. At front centre is TH5 (running to the left) and above it are helices TH6 and TH7. Behind these pairs of antiparallel helices is another pair of antiparallel helices, TH8 and TH9, with TH9 running upwards and ending at residue 378. *a*, Asp and Glu side chains are shown. Notice the tips of two helix layers, TL3 and TL5 contain a total of six acidic groups (on the left). *b*, Lys, Arg and His side chains are shown. Notice the positive charge asymmetry, with all charges at the bottom and back of the domain, except Lys 299 near loop TL3 between TH5 and TH6.

of RB1 and RB2. The other interfaces are not common among three different crystal forms. The inability of the dimer to bind to the DT receptor¹⁶ suggests that the dimer interaction sterically blocks the receptor-binding domains of each monomer from the receptors on the surface of a target cell. The conformational differences between the monomer in the dimer and the native monomeric DT remain uncertain, but biochemical evidence suggests they are not large. Binding data show the affinity constant of the dimer for ApUp is the same as that of the monomer, and that the dimer binds two ApUps¹⁶. In addition, comparable specific activities of NAD-glycohydrolase activity and affinities for NAD^+ were found in the monomer and dimer; and the specific ADP-ribosyltransferase activity of

fragment A released from the dimer after reduction was the same as that from the monomer¹⁶. These findings show that the conformation of the C domain, and of the portion of the R domain interfacing the C domain, are relatively unperturbed in the dimer.

Chimaeric toxins. The three-dimensional structure of DT demonstrates that the receptor-binding function is associated with a discrete, compact domain and defines the boundary between the R domain and the C and T domains. This finding may permit the design of re-engineered DT molecules in which the R domain is replaced by immunoglobulin or other domains that target specific cell types for killing by chimaeric toxins. Construction of a hybrid protein using gene splicing has been

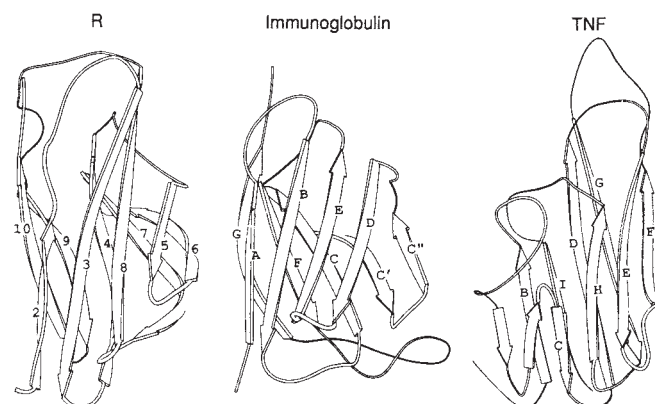


FIG. 4 The topology of the R domain of DT (left) is compared to that of an immunoglobulin variable domain (centre)⁴⁷ and to tumour necrosis factor (TNF, right)⁴⁸. R domain is viewed in the direction from the back of DT in Fig. 1. Numbers from 2 to 10 of the R domain represent the strands RB2 to RB10 of DT. Notice that strands 2, 3, 4, 8, 9 and 10 of the R domain correspond well to strands A, B, C, E, F and G of the immunoglobulin variable domain. Also, strands 3, 4, 5, 6, 7, 8 and 9 correspond well to strands C, D, E, F, G, H and I of TNF, a classic jelly-roll⁴⁹.

reported⁵⁰, where the R domain was nearly replaced by an interleukin-2. Elucidation of the boundary between the C and

T domains and the R domain may now facilitate the design of more effective chimaeras of DT. □

Received 17 February; accepted 30 March 1992.

- Greenfield, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 6853-6857 (1983).
- Moskaug, J. O., Sletten, K., Sandvig, K. & Olsnes, S. *J. biol. Chem.* **264**, 15709-15713 (1989).
- Collier, R. J. & Kandel, J. *J. biol. Chem.* **246**, 1496-1503 (1971).
- Sandvig, K. & Olsnes, S. *J. biol. Chem.* **256**, 9068-9076 (1981).
- Yamaizumi, M., Mekada, E., Uchida, T. & Okada, Y. *Cell* **15**, 245-250 (1978).
- Colombatti, M., Dell'Arciprete, L., Rappuoli, R. & Tridente, G. *Meth. Enzym.* **178**, 404-422 (1989).
- Williams, D. P. *et al. Prot. Engng* **1**, 493-498 (1987).
- Chaudhary, V. K., Gallo, M. G., Fitzgerald, D. J. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9491-9494 (1990).
- Greenfield, L., Johnson, V. G. & Youle, R. J. *Science* **238**, 536-539 (1987).
- Papini, E. *et al. Eur. J. Biochem.* **169**, 637-644 (1987).
- Moskaug, J. O., Stenmark, H. & Olsnes, S. *J. biol. Chem.* **266**, 2652-2659 (1991).
- Blewitt, M. G., Chung, L. A. & London, E. *Biochemistry* **24**, 5458-5464 (1985).
- Kagan, B. L., Finkelstein, A. & Colombini, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4950-4954 (1981).
- Donovan, J. J., Simon, M. I., Draper, R. K. & Montal, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 172-176 (1981).
- Collier, R. J., Westbrook, E. M., McKay, D. B. & Eisenberg, D. *J. biol. Chem.* **257**, 5283-5285 (1982).
- Carroll, S. F., Barbieri, J. T. & Collier, R. J. *Biochemistry* **25**, 2425-2430 (1986).
- Fujii, G., Choe, S., Bennett, M. J. & Eisenberg, D. *J. molec. Biol.* **222**, 861-864 (1991).
- Wang, B.-C. *Meth. Enzym.* **115**, 90-112 (1985).
- Brünger, A. T. *Acta crystallogr.* **A47**, 195-204 (1991).
- Rossmann, M. G. & Blow, D. M. *Acta crystallogr.* **15**, 24-31 (1962).
- Zhang, K. Y. J. & Main, P. *Acta crystallogr.* **A46**, 377-381 (1991).
- Lüthy, R., Bowie, J. U. & Eisenberg, D. *Nature* **356**, 83-85 (1992).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110-119 (1991).
- Brünger, A. T. *Nature* **355**, 472-475 (1992).
- Allured, V. S., Collier, R. J., Carroll, S. F. & McKay, D. B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1320-1324 (1986).

- Carroll, S. F. & Collier, R. J. *Molec. Microbiol.* **2**, 293-296 (1988).
- Brandhuber, B. J., Allured, V. S., Falbel, T. G. & McKay, D. B. *Proteins* **3**, 146-154 (1988).
- Sixma, T. K. *et al. Nature* **351**, 371-377 (1991).
- Parker, M. W., Pattus, F., Tucker, A. D. & Tsernoglou, D. *Nature* **337**, 93-96 (1989).
- Richardson, J. *Adv. Prot. Chem.* **34**, 167-339 (1981).
- Li, J. D., Carroll, J. & Ellar, D. J. *Nature* **353**, 815-821 (1991).
- Carroll, S. F. & Collier, R. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3307-3311 (1984).
- Papini, E., Schiavo, G., Sandona, D., Rappuoli, R. & Montecucco, C. *J. biol. Chem.* **264**, 12385-12388 (1989).
- Papini, E. *et al. J. biol. Chem.* **266**, 2494-2498 (1991).
- Giannini, G., Rappuoli, R. & Ratti, G. *Nucleic Acids Res.* **12**, 4063-4069 (1984).
- Collins, C. M. & Collier, R. J. *Biochim. biophys. Acta* **828**, 138-143 (1985).
- Zhao, J.-M. & London, E. *Biochemistry* **27**, 3398-3403 (1988).
- Proia, R. L., Wray, S. K., Hart, D. A. & Eidsels, L. *J. biol. Chem.* **255**, 12025-12033 (1980).
- Lory, S. & Collier, R. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 267-271 (1980).
- Rees, D. C., DeAntonio, L. & Eisenberg, D. *Science* **245**, 510-513 (1989).
- McLaughlin, S. *Curr. Top. Memb. Trans.* **9**, 71-144 (1977).
- Weissman, L. thesis, Univ. California (1979).
- Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. *Acta crystallogr.* **A43**, 1-5 (1987).
- Jones, A. T. *Meth. Enzym.* **115**, 157-171 (1985).
- Ponder, J. W. & Richards, F. M. *J. molec. Biol.* **193**, 775-791 (1987).
- Brünger, A. T. & Krukowski, A. *Acta crystallogr.* **A46**, 585-593 (1990).
- Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. *J. molec. Biol.* **141**, 369-391 (1980).
- Eck, M. J. & Sprang, S. R. *J. biol. Chem.* **264**, 17595-17605 (1989).
- Jones, E. Y., Stuart, D. I. & Walker, N. P. C. *Nature* **338**, 225-228 (1989).
- Kiyokawa, T., Williams, D. P., Snider, C. E., Strom, T. B. & Murphy, J. R. *Prot. Engng* **4**, 463-468 (1991).

ACKNOWLEDGEMENTS. We thank E. M. Westbrook, B. Dijkstra, S. F. Carroll and J. T. Barbieri for much work and D. B. McKay for providing the coordinates of ETA, the NIH for support, and a traineeship for MJB

LETTERS TO NATURE

Discovery of soft X-ray pulsations from the γ -ray source Geminga

J. P. Halpern* & S. S. Holt†

* Columbia Astrophysics Laboratory, Columbia University, 538 West 120th Street, New York, New York 10027, USA
† Director of Space Sciences, Code 600, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

THE nature of the γ -ray source 'Geminga' (2CG195+04) is a problem of considerable importance in high-energy astrophysics. First discovered in 1972 by the SAS-2 satellite¹, Geminga emits virtually all its power at energies above 50 MeV, and at energies above 100 MeV is the second brightest source in the γ -ray sky survey made by the Cos-B satellite². It eluded identification at all other wavelengths until the Einstein Observatory found an unusual soft X-ray source, 1E0630+178, in its error box³. This source also has a claimed twenty-fifth magnitude optical counterpart⁴⁻⁶. This distinctive set of properties is reminiscent of the Vela pulsar, except for the absence of radio emission⁷ or a synchrotron nebula³. We have made a more sensitive soft X-ray observation of the Geminga field using Rosat, and have detected coherent pulsations from 1E0630+178 at a period of 0.237 s. This result confirms suggestions^{3-6,8,9} that Geminga is, like Vela, a γ -ray pulsar. We speculate that Geminga is somewhat the older of the two. With this discovery we consider the mystery of Geminga largely solved.

For Geminga, the soft X-ray band is the most promising one for a pulsar search because the count rate is high and the background is negligible. The Geminga field was observed with the Position Sensitive Proportional Counter (PSPC) aboard Rosat¹⁰ during 1991 March 14-17. A total of 14,390 s of exposure were obtained during 10 satellite orbits as listed in Table 1. The mean count rate of 1E0630+178 was 0.53 s⁻¹ in the 0.07-2.4-keV band. A total of 7,630 counts were detected within a circle of radius 1.5' about the source centroid, of which only ~1% are background. After transforming the photon arrival times to the barycentre of the Solar System using the accurate (~3") Einstein High Resolution Imager position, the search for periodicity was

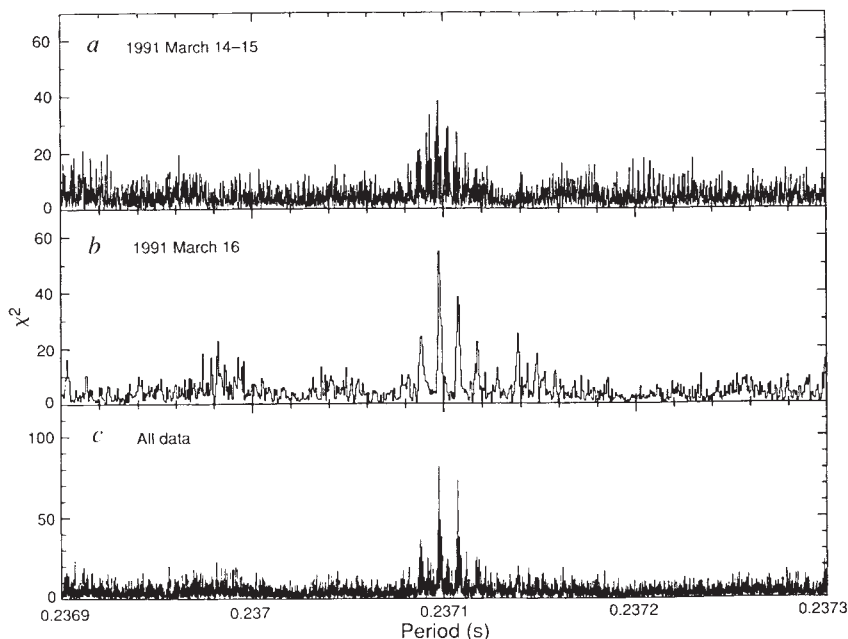
TABLE 1 Rosat observations of Geminga

Date (UT)	Start time (s)	Duration (s)	Counts
1991 March 14	76,564	1,193	612
1991 March 15	1,191	1,437	749
1991 March 15	29,362	1,360	703
1991 March 15	64,656	1,567	866
1991 March 15	76,468	1,272	641
1991 March 16	64,514	1,690	887
1991 March 16	69,999	1,959	1,038
1991 March 16	76,448	1,272	724
1991 March 16	81,745	1,509	786
1991 March 17	46,703	1,131	624

begun by performing a fast Fourier transform (FFT) on the section of data consisting of four consecutive orbits on March 16. The FFT covered 262,144 independent trial periods down to a minimum period of 0.072 s. The only significant peak in the power spectrum occurred at a period of 0.237097 s, for which the power is 17.5 times the mean. The probability that a peak of this strength or greater will occur by chance somewhere in the power spectrum is 0.007. The same section of data was then folded about a range of trial periods. The χ^2 values of the resulting five-bin folded light curves were calculated with respect to a constant mean. The candidate period appeared with $\chi^2 = 39.5$ for 4 degrees of freedom. The probability that χ^2 greater than or equal to this value will occur by chance in a single trial is 6×10^{-8} . When multiplied by the number of trial periods, the chance probability is 0.016. The highest χ^2 in the periodogram of the folded data is 54.7 (Fig. 1b). The single-trial chance probability is less than 10^{-10} , but the effective number of periods searched in the epoch folding is about a factor of 5 higher than in the FFT. We conclude that the probability that the period is spurious is $\leq 1 \times 10^{-4}$.

The reality of the signal at 0.237 s is further supported by its presence in the remaining half of the data which was not included in the original FFT. Figure 1a shows the periodogram for the five orbits of data obtained on March 14-15. The period is clearly indicated in this independent set of data, although the less-than-optimal sampling introduces many more aliases at the 96-min satellite orbital period and its multiples. Figure 1c is the result of the coherent folding of data from all 10 orbits. The

FIG. 1 Periodograms of the five-bin folded light curves. The multiple peaks are aliases of the 96-min satellite orbital period. The best-fit period is $0.2370974 \text{ s} \pm 0.1 \text{ } \mu\text{s}$.



highest peaks in all three panels are at the same period to within the errors. The symmetry of the peak structure indicates that the highest peak at $0.2370974 \text{ s} \pm 0.1 \text{ } \mu\text{s}$ is most likely to be the true period. In addition to these analyses, the period is detected weakly in those individual orbits having more than 1,500 s of exposure. There is no evidence for orbital motion or a secular change in period P during the span of these observations. On a timescale of two days, the upper limit on $\dot{P}$ is $2 \times 10^{-12} \text{ s s}^{-1}$, and Δv is less than 0.5 km s^{-1} . We conclude that the 0.237-s period is not an instrumental artefact, because similar analyses on background counts collected at different positions in the detector do not reveal a signal. We do not detect a significant signal at any other period, including previously reported periods^{11–13} near 60 s, even though the PSPC is at risk of manufacturing spurious periods between 30 and 400 s because of the deliberate dithering of the satellite.

Figure 2 shows the folded light curve in three different energy bands, and their sum. The pulse profile has a single broad peak, which is consistent with the absence of power at submultiples of the 0.237-s period. The pulsed fraction is 24% in the band 0.07–0.2 keV and declines with energy. At 0.2–0.28 keV, the pulsed fraction is 19%, and at 0.28–1.50 keV it is 15%. These pulsed fractions are probably lower limits, as there is a known timing error in the processed data which shifts some of the photon arrival times randomly with respect to the phase of the 0.237-s period. The extremely soft spectrum of the pulsations, and the broadness of the pulse profile, are suggestive of thermal emission from the surface of a neutron star. But it is difficult to interpret these data in terms of surface temperature and distance, because the X-ray spectrum in the Rosat bandpass can be modified from the black-body shape as a result of non-grey atmospheric opacity¹⁴. Strong L and K edges could be present. Rosat has recently detected soft X-ray pulsations from Vela¹⁵, also in the lowest energy channels.

To study the X-ray spectrum, we extracted all the source counts within a $2'$ radius, and subtracted background from an annulus between radii of $2.5'$ and $4.2'$. The total flux in the 0.1–2.4-keV band is $\sim 1.5 \times 10^{-12} \text{ erg cm}^{-2} \text{ s}^{-1}$, which is roughly

consistent with the Einstein result. The X-ray spectrum is very soft, but it cannot be adequately fitted by a black-body or a power-law spectrum, as illustrated in Fig. 3. We have not yet explored the full range of plausible spectral models, but with the caveat concerning atmospheric opacity in mind, a composite model consisting of a black body of temperature $T \approx (3-4) \times 10^5 \text{ K}$ plus a power-law component with energy index α in the range 0.75–1.75 yields an adequate fit to the data. In

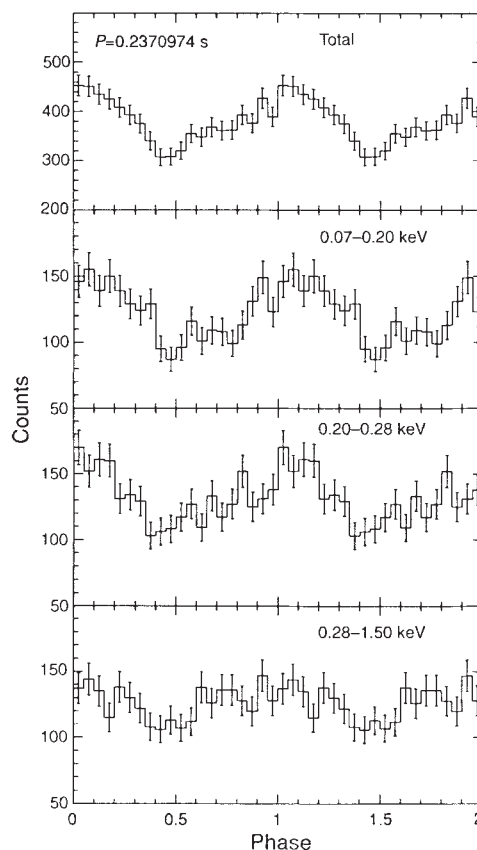
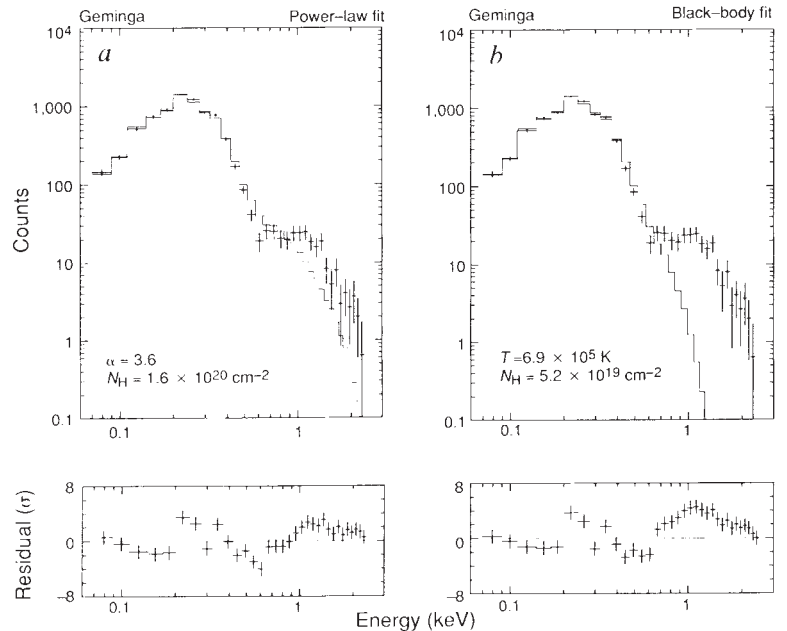


FIG. 2 Pulse profile of all data between 0.07 and 1.5 keV folded into a 20-bin light curve. The top panel is the sum of the bottom three.

FIG. 3 Spectral fits with single-component models. *a*, Power law; *b*, black body.



this model, the decrease of pulsed fraction with energy might be caused by the increasing contribution of an unpulsed synchrotron component. A column density in the range $(0.5\text{--}3) \times 10^{20} \text{ cm}^{-2}$ is allowed.

Observations indicating that Geminga might be a pulsar somewhat older than Vela have been reviewed previously^{3-6,8}. The overall energy distributions of the two objects are similar at wavelengths from optical through γ -rays. The ratio of $L_\gamma/L_x = 1,000$ is a likely indicator of a pulsar emission mechanism. The ratio $L_x/L_{\text{opt}} \approx 1,800$, and the soft X-ray spectrum, also suggest an isolated neutron star. The discovery of a short pulse period fulfils this expectation. A detailed theoretical model for γ -ray pulsars was developed by Ruderman and Cheng⁹, who proposed an evolutionary scheme in which a short-period γ -ray pulsar evolves from a Crab-like stage, through that of Vela, and finally reaches that of Geminga. In the regime of Vela-like pulsars, γ -rays are produced in the outer magnetosphere as synchrotron radiation from electron-positron pairs. As it slows down, such a pulsar uses an ever increasing fraction of its spin-down power in the production of an e^+ wind and associated γ -ray emission until almost the full spin-down power, $\sim 4 \times 10^{36} \text{ erg s}^{-1}$, goes into this mechanism. As the period increases further, the currents necessary to sustain pair production are quenched and the γ -ray emission ceases. They predicted that Geminga is near the turnoff period, which for Vela itself would be $\sim 0.13 \text{ s}$, and postulated that ordinary pulsars avoid this fate because they have more highly curved magnetic field lines near the polar cap, which can maintain the potential drop required for pair production.

The theory was generalized by Chen and Ruderman¹⁶, who argued that there is actually a death line for Vela-like (outer gap) pulsars given by $5 \log B - 12 \log P = 72.2$, where B is the surface magnetic field. If Geminga is to fall near this line, it must have $B \approx 1 \times 10^{13} \text{ G}$, only a factor of 3 greater than that of Vela. In this case, the total spin-down power, $\dot{E} \approx B^2 R^6 \Omega^4 / c^3$, would be $\sim 1.8 \times 10^{36} \text{ erg s}^{-1}$. As the observed γ -ray flux¹ is at least $2.4 \times 10^{-9} \text{ erg cm}^{-2} \text{ s}^{-1}$, an upper limit to the distance is $2,500 (B/10^{13}) \text{ pc}$ for the luminosity not to exceed the estimated spin-down power. (The column density derived from the X-ray spectrum almost certainly restricts the distance to less than 500 pc .) The predicted magnetic field would result in a spin-down rate of $4 \times 10^{-13} (B/10^{13})^2 \text{ s s}^{-1}$, significantly below the upper limit of 2×10^{-12} established by this observation. A second observation could easily determine $\dot{P}$, and thus the magnetic field. The fact that Geminga is not detected in high-energy X-rays

or low-energy γ -rays¹⁷ is consistent with the outer gap model. The low-energy cutoff of the spectrum at a frequency ω_c is determined by the energy of the e^+ pairs that leave the magnetosphere before losing their energy to synchrotron radiation. Because $\omega_c \propto \Omega^{-7} B^{-3}$, scaling from the Vela break at $\sim 1 \text{ MeV}$ (ref. 7) predicts a break at $\sim 25 \text{ MeV}$ for Geminga. Other pulsars that may be progenitors of the Geminga-like γ -ray pulsars are PSR1509-58 (ref. 18) with $P = 0.150 \text{ s}$ and PSR1706-44 with $P = 0.102 \text{ s}$. These both fall in the region of Vela-like pulsars¹⁶, and were recently detected^{19,20} by the Compton Gamma Ray Observatory (CGRO).

Our results thus largely confirm the expectation that Geminga is a pulsar similar to Vela. The absence of a synchrotron nebula around Geminga remains a fundamental difference, which is perhaps related to the fact that Geminga has the longest period of the γ -ray pulsars. We suggest that many of the other unidentified high-energy γ -ray sources are spinning neutron stars which are 'radio quiet', or whose radio beams do not intersect the Earth. Rosat should be able to identify γ -ray pulsars by finding the periods of sources detected by CGRO in the galactic plane. $\square$

Received 30 March; accepted 28 April 1992.

1. Fichtel, C. E. *et al. Astrophys. J.* **198**, 163-182 (1975).
2. Swanenburg, B. N. *et al. Astrophys. J.* **243**, L69-L73 (1981).
3. Bignami, G. F., Caraveo, P. A. & Lamb, R. C. *Astrophys. J.* **272**, L9-L13 (1983).
4. Bignami, G. F. *et al. Astrophys. J.* **319**, 358 (1987).
5. Halpern, J. P. & Tytler, D. *Astrophys. J.* **330**, 201-217 (1988).
6. Bignami, G. F., Caraveo, P. A. & Paul, J. A. *Astr. Astrophys.* **202**, L1 (1988).
7. Spoelstra, T. A. & Hermesen, W. *Astr. Astrophys.* **135**, 135-140 (1984).
8. Halpern, J. P. in *Proc. GRO Science Workshop* (ed Johnson, N.) 4-166-4-173 (1989).
9. Ruderman, M. & Cheng, K. S. *Astrophys. J.* **335**, 306-318 (1988).
10. Trümper, J. *et al. Adv. Space Res.* **2**, 241-249 (1983).
11. Bignami, G. F., Caraveo, P. A. & Paul, J. A. *Nature* **310**, 464-469 (1984).
12. Bignami, G. F. in *The Origin and Evolution of Neutron Stars* (eds Helfand, D. J. & Huang, J. H.) 465-474 (Reidel, Dordrecht, 1987).
13. Caraveo, P. A. & Bignami, G. F. in *The Origin and Evolution of Neutron Stars* 545 (eds Helfand, D. J. & Huang, J. H.) (Reidel, Dordrecht, 1987).
14. Romani, R. W. *Astrophys. J.* **313**, 718-726 (1987).
15. Ögelman, H., Finley, J. P., Aschenbach, B., Trümper, J. & Zimmermann, U. *Bull. Am. astr. Soc.* **23**, 1349 (1991).
16. Chen, K. & Ruderman, M. *Astrophys. J.* (in the press).
17. Bignami, G. F. & Hermesen, W. A. *Rev. Astr. Astrophys.* **21**, 67-108 (1983).
18. Seward, F. D. & Harnden, F. R. H. Jr. *Astrophys. J.* **256**, L45-L47 (1982).
19. Wilson, R. B., Finger, M. H., Fishman, G. J., Meegan, C. A. & Paciesas, W. S. *IAU Circ. No.* 5429 (1992).
20. Kniffen, D. A. *et al. IAU Circ. No.* 5485 (1992).

ACKNOWLEDGEMENTS. We thank F. R. Harnden and F. Seward for making the barycentric correction, and for assistance with the data analysis. We thank G. Bignami for his review. This work was supported by NASA through a Rosat guest observer grant.

Instability and reconnection in the head-on collision of two vortex rings

T. T. Lim & T. B. Nickels

Department of Mechanical & Manufacturing Engineering,
University of Melbourne, Parkville 3052, Australia

ONE mechanism by which fluid flows increase their complexity is through the instability of vortex filaments. When an instability brings vortex filaments of opposite circulation together, the filaments may break and rejoin in a process known as reconnection. This process of instability and reconnection leads to some fundamental changes in the topology of flows. Here we present experimental observations of a special type of instability in which two colliding vortex rings become unstable and reconnect to form a series of smaller rings. Although this phenomenon was briefly noted more than a decade ago¹, no detailed observations were made, and little is known about the mechanisms involved. We have used coloured dyes to reveal the detailed structure of the small rings and many other features, including a short-wavelength instability around the circumference of the colliding rings. At high Reynolds number, collision leads to a turbulent cloud, with the occasional appearance of small rings.

The experiment was conducted in water in a glass tank (1.22 m long, 0.36 m wide and 0.47 m deep) in which were immersed two horizontally opposed nozzles spaced 220 mm apart. The position of one nozzle could be finely adjusted to make the rings collide exactly head-on. Both nozzles were connected to a piston, which ejected short, equal pulses of water from both nozzles simultaneously to produce two identical vortex rings travelling towards each other. Accurate, repeatable results were achieved

by driving the piston with an electronically controlled stepping motor: this meant that the circulation and Reynolds number of the rings could also be determined. The vortex rings were made visible by releasing neutrally buoyant dyes around the circumference of each nozzle; the resulting flow patterns were recorded using a video recorder.

Figure 1 shows different stages of a head-on collision for a Reynolds number (Re) of $\sim 1,000$. (The initial Reynolds number of each ring is defined by UD/ν , where U is the initial translation velocity, D is the diameter of the ring and ν is the kinematic viscosity.) Figure 1*b* shows that when the two vortex rings are close to one another, the velocity induced by one ring on the other causes both rings to grow in diameter. The early stages of this growth follow the predictions of an inviscid analysis reasonably well, but when each ring has increased in size to about four times its initial diameter, a symmetrical instability in the form of azimuthal waviness begins to develop. As time progresses, the waves on the rings grow until they touch at the locations of maximum inward displacement. At the points of contact, the segments of the two vortex filaments eventually become interconnected to form small rings, a process commonly referred to as 'vortex reconnection'. As can be seen in Fig. 1*e*, the observation that each small ring is made up of both red and blue dye indicates that it consists of segments from both of the original rings (Fig. 2 shows a close-up view of the small rings). Throughout the process of reconnection, the original vortex rings continue to grow in diameter, albeit at a slower rate. This growth is associated with stretching of the contact regions between the waves, and seems to be related to the reconnection process. Once the small rings have fully formed, the original rings cease to exist, and the small rings then convect away radially from the central axis at slightly different speeds. The azimuthal waves that occur during the collision do not always form a regular pattern around the rings: the wavelength of the instability varies along the circumference and from run to run.

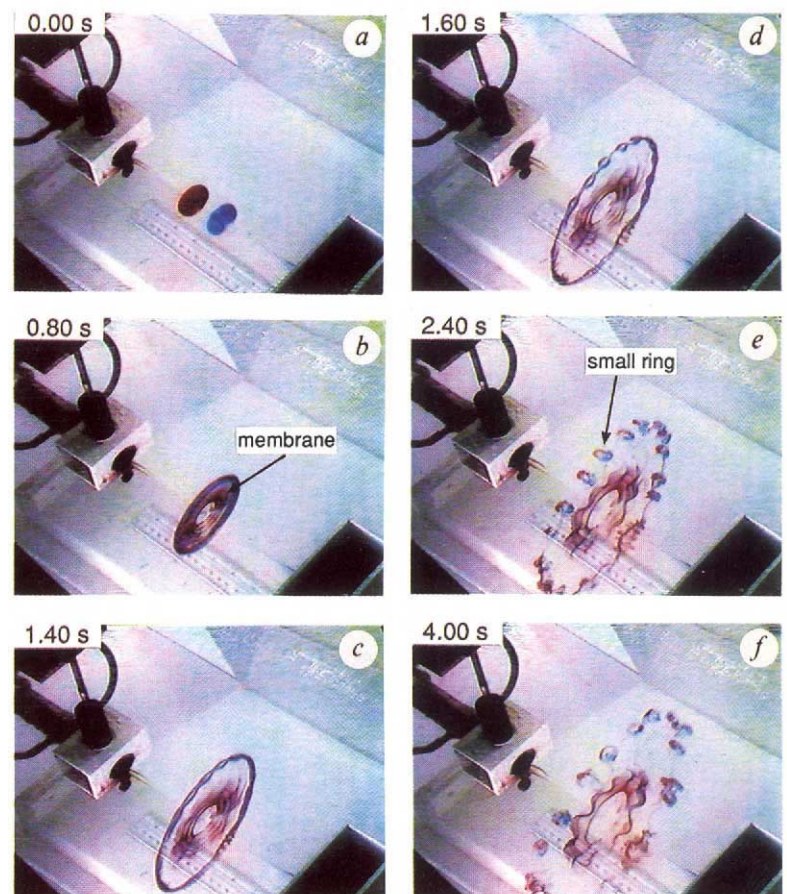


FIG. 1 A sequence of photographs showing different stages of the head-on collision between two identical vortex rings. The initial Reynolds number of each ring is roughly 1,000. The first photograph in the sequence, *a*, has been arbitrarily assigned as $t = 0.00$ s. The elapsed time for the subsequent stages of the collision is shown in each photograph, with different time intervals adopted to illustrate the main features of the flow. The 'membrane' structure with concentric ribs observed in *b*–*f* occurs because the vortex rings consist of rolled-up spiral dye sheets which become squashed and flattened during the collision. Each turn of the sheet forms a fold which becomes one of the concentric ribs on the membrane. Because of the effect of viscous diffusion, we do not believe that the membrane contains much vorticity.

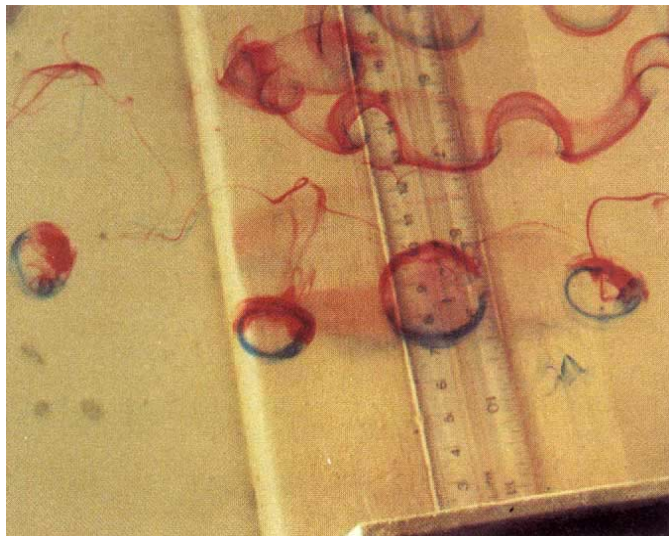


FIG. 2 Close-up view of the small rings.

In some cases the waves did not develop significantly over a considerable part of the circumference. Small rings have also been observed to form during the collision of offset vortex rings².

Unexpectedly, a short-wavelength instability also formed in some of the runs (see Fig. 3). To our knowledge, this is the first time that this type of instability has been identified in colliding

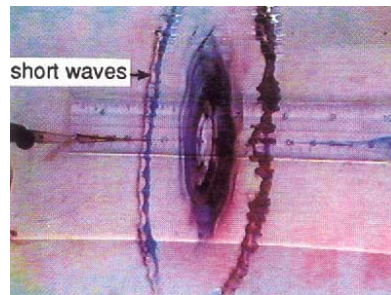


FIG. 3 Short-wavelength instability. This instability can also be seen on part of the rings' circumference in Fig. 1c and d.

vortex rings. On some occasions, both modes of instability coexisted on the same ring and did not interact noticeably (see Fig. 1c, d). The short-wavelength instability is similar in appearance to the 'core bulging' or 'bursting' instability observed in trailing vortices, which numerical analysis³ suggests is caused by the effect of viscosity.

Figure 4 shows similar collisions at higher Reynolds numbers. For a Reynolds number between 1,000 and 1,500 (Fig. 4a), the general behaviour of the vortex rings is similar to that shown in Fig. 1, although the wavelength of the long-wave instability seems to have decreased. Furthermore, the short-wavelength instability discussed earlier did not occur in this case: this is consistent with the suggestion that the instability is viscous in nature, as at higher Reynolds number the effect of viscous forces

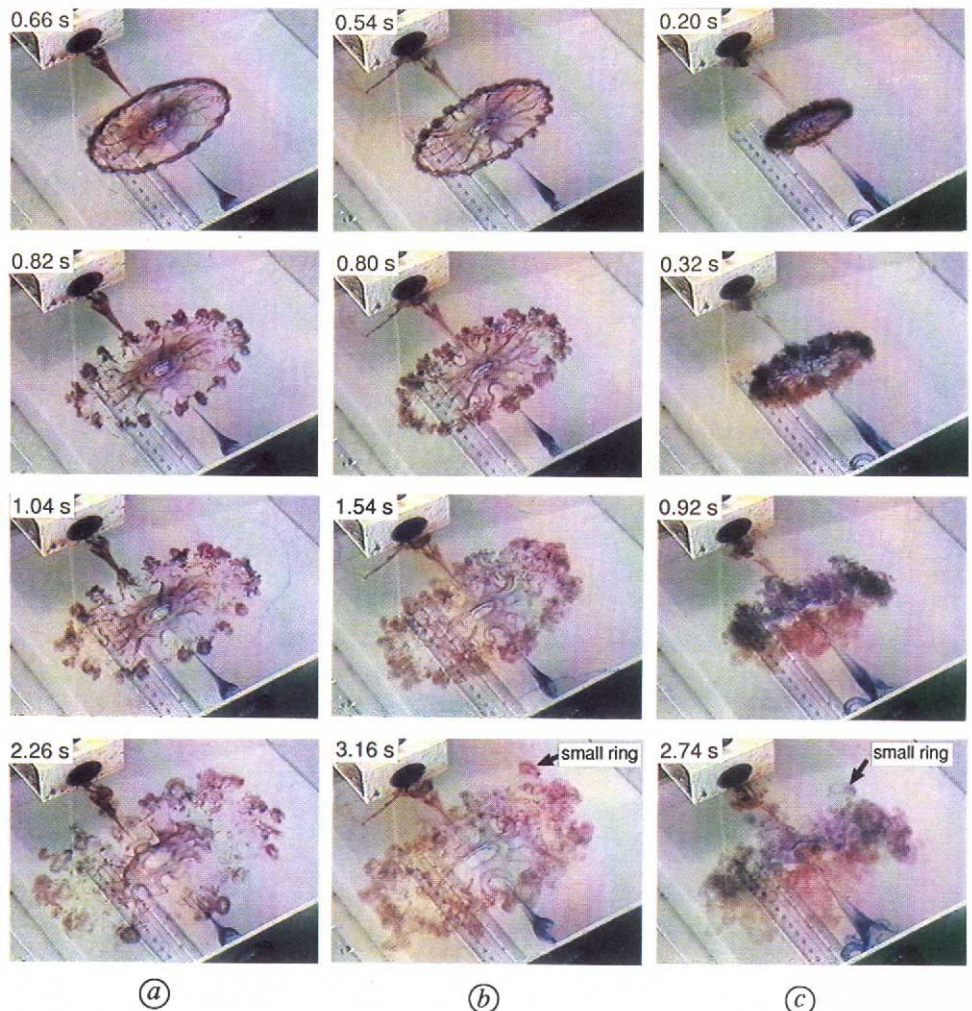


FIG. 4 Collision at higher Reynolds numbers. a, $Re \approx 1,450$; b, $Re \approx 1,850$; c, $Re \approx 3,500$. The position of the rings at time $t = 0.00$ s is the same as in Fig. 1a.

becomes less important. As the Reynolds number is increased from $\sim 1,500$ to $3,000$ (Fig. 4b), the flow goes through a transition in which the small rings appear to have fine structures superimposed on them. When the Reynolds number is increased beyond $3,000$, the two rings disintegrate into a turbulent cloud as shown in Fig. 4c. This disintegration has also been observed by Schultz-Grunow⁴. In the two cases at higher Reynolds number, a single small ring may be seen escaping from the interaction region, even though the rest of the flows are complicated (see the final photographs in Fig. 4b, c). It may be that in some places around the circumference, the local conditions are more characteristic of a lower Reynolds number.

The development of the azimuthal waves and the eventual formation of the small rings is reminiscent of the interaction between the two trailing vortices produced by the wing tips of an aircraft (ref. 5). Using an inviscid model, Crow⁶ analysed the growth of waves on the trailing vortices in order to predict the wavelength that would dominate. The similarity between the two flow cases and the noted inviscid nature of the early stages of the interaction prompted us to attempt to apply this model. In this approach, the size and separation of the cores determine the wavelength of the instability, which increases with increasing core size and decreasing separation (although the dependence on the core size is weak). In this case, the core size is estimated using the circulation and measured translational velocity, and the core separation determined from a video of laser cross-sections of the flow (not shown here). In Crow's model there is no stretching of the filaments and the wavelength of the instability does not change with time. In our case, the wavelength of the instability increases as the rings stretch, although the number of waves on the circumference remains constant. Thus as a first approximation, we ignore the stretching and consider that the instability forms at the instant when the azimuthal waves first become apparent on the circumference of the rings. From the values of the core diameter and separation at that instant, the wavelength of the instability can be determined using Crow's model, and the number of small rings that should form can be estimated by dividing the circumference of the vortex rings by the calculated wavelength. Given that the ratio of the radius of curvature to the core separation is ~ 50 , the curvature of the filaments can be neglected and they are treated as a simple trailing vortex pair. Within the accuracy of the measurements, the predicted number of rings (N) falls between 11 and 24 for all Reynolds numbers in the range 860 – $1,500$, in rough agreement with the experimental observations (on average, N increases from 15 to 20 as Re increase from 860 to $1,500$). Although it is difficult to measure the circulation and core separation very accurately, the agreement of this first approximation with observation suggests that further investigation may be worthwhile.

It is important to note that the analysis of Crow⁶ breaks down when the filaments touch, and can therefore only explain the initial growth of the waves and not the formation of the rings. The rings are formed by the breaking and rejoining of the vortex filaments, a process commonly referred to as 'reconnection'. This has been the subject of much recent study, and although not all researchers agree as to the detailed mechanism involved, the basic overall processes they discuss are similar. Briefly, when two vortex filaments are brought into contact by an induced velocity field, viscous diffusion in the contact region causes annihilation of vorticity. The annihilation of vorticity effectively 'severs' the filaments, and, due to the kinematic constraint that vortex lines cannot end inside a fluid, they reconnect on either side of the contact region. For a summary of different views of the detailed mechanisms involved, see refs 7–10. □

4. Schultz-Grunow, F. *Flow Visualization II* (ed. Merzkirch, W.) 361–365 (Hemisphere, New York, 1982).
5. Bisgood, P. L. *Royal Aircraft Establishment TM FS no. 330* (1980).
6. Crow, S. C. *Am. Inst. Aeronautics Astronautics J.* **8**, 2172–2179 (1970).
7. Saffman, P. G. *J. Fluid Mech.* **212**, 395–402 (1990).
8. Melander, M. V. & Hussain, F. *Phys. Fluids A1*, 633–636 (1989).
9. Kida, S., Takaoka, M. & Hussain, F. *Phys. Fluids A1*, 630–632 (1989).
10. Zawadzki, I. & Aref, H. *Phys. Fluids A3*, 1405–1410 (1991).

ACKNOWLEDGEMENTS. We thank A. E. Perry for his encouragement and support and D. H. Thompson for assistance with the reproduction of the colour photographs. This work was supported by the Australian Research Council.

Coherence of the superconducting wavefunction between the heavy-fermion superconductor UPd₂Al₃ and niobium

Y. He[†]*, C. Muirhead^{*}, A. Bradshaw[‡], J. S. Abell[‡],
C. Schank[§], G. Geibel[§] & F. Steglich[§]

^{*} School of Physics and Space Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

[‡] School of Metallurgy and Materials, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

[§] SFB 252, Institut für Festkörperphysik, Technische Hochschule Darmstadt, W6100 Darmstadt, Germany

HEAVY-fermion superconductors (in which the charge carriers have an effective mass ~ 100 times the free electron mass) have been the subject of intense study during the past decade¹, in part because of the suggestion that some of these materials may exhibit non-conventional pairing mechanisms². Here we report a demonstration of quantum coherence of the superconducting wavefunction between the conventional superconductor niobium and the recently discovered³ heavy-fermion superconductor UPd₂Al₃, which is a potential candidate for a non-conventional pairing mechanism⁴. The experimental method is similar to our earlier work on high- T_c materials^{5,6}: we use a small pointed rod of UPd₂Al₃ to bridge the gap in an almost closed niobium ring. We observe persistent currents in the composite ring, and trapped flux, which is in discrete quantum states separated by the flux quantum $h/2e$. Although the observation of phase coherence between UPd₂Al₃ and niobium may not constrain the nature of the pairing in UPd₂Al₃, we also observe Josephson-like current-voltage characteristics at the junction, but with very small products of critical current and normal-state resistance. If other possible causes can be eliminated, these small $I_c R_n$ values may point to a small tunnelling probability between the two superconductors, and hence to unconventional pairing in UPd₂Al₃.

In a conventional BCS superconductor the electrons are paired in a spin singlet s -state. It is in principle possible to have other forms of pairing (p or d), and there has been much discussion as to the possible difficulty in obtaining coherent tunnelling of Cooper pairs between systems with superconducting states of different symmetry (see for example refs 6–8). It is now generally believed that the symmetry-breaking effect of the barrier allows coupling between states of different orbital angular momentum, and that coherent tunnelling between states with singlet and triplet pairing is also possible, albeit with reduced matrix elements⁷. The possibility of superconducting states involving quasiparticles with even more exotic symmetry, such as anyons, has recently excited interest⁹, but the consequences for coupling across a barrier between anyonic and non-anyonic superconductors are unclear. In any case, demonstration of coherence in systems having the potential for mixed symmetry is of fundamental importance.

[†] On leave from Tsinghua University, Beijing 100084, China.

Received 4 December 1991; accepted 17 March 1992.

1. Oshima, Y. *J. phys. Soc. Japan* **44**, 328–331 (1978).

2. Smith, G. B. & Wei, T. *Bull. Am. phys. Soc. HG 5* **36**, 2694 (1991).

3. Khorrami, M. R. *J. Fluid Mech.* **225**, 197–212 (1991).

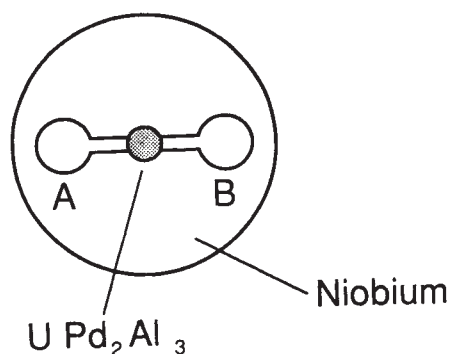


FIG. 1 Schematic of the experiment. The UPd_2Al_3 pin is lowered into the slot in the niobium to form superconducting weak links at the contact points.

Convincing tunnelling experiments between conventional and heavy-fermion superconductors are difficult, as one must eliminate the possibility that a small amount of one superconductor has wiped off on the other and that one is simply observing tunnelling between identical materials. Suppression of superconductivity by temperature in one component and by magnetic field in the other is one method that has been demonstrated in the $\text{CeCu}_{2.2}\text{Si}_2/\text{Al}$ system¹⁰. The demonstration of flux quantization, on the other hand, is unambiguous, as it relies solely on the requirements that the superconducting wavefunction retains coherence in going once around the ring and that the wavefunction is single-valued at a point.

Several of the heavy-fermion materials (such as UPt_3 , $\text{U}_{1-x}\text{Th}_x\text{Be}_{13}$ (ref. 7) and the subject of the present study, UPd_2Al_3 (refs 3, 4)) are considered to be potential candidates for non-conventional pairing. The first two have transition temperatures below 1 K. The latter is superconducting below 2 K, and experiments can easily be done in a simple pumped ^4He cryostat.

The experimental arrangement is shown in Fig. 1. The niobium sample is in the form of a disc 3 mm thick with two 3-mm-diameter holes. The holes are connected by a 0.3-mm-wide slot. The polycrystalline sample of UPd_2Al_3 was prepared as described in ref. 4.

A bar 4 mm long and 2.5 mm square was cut by spark erosion, and a cone with an opening angle of 90° was formed on one end, also by spark erosion, while the rod was rotated at a few revolutions per second. The rod was encapsulated in a brass holder, which could be forced into the slot in the niobium with a screw to make the required superconducting junctions. The drive mechanism was always withdrawn after a suitable contact had been made, to avoid problems with vibration or thermal contraction during the measurements. A small, known magnetic flux could be applied to the system by means of a long thin

solenoid passing through hole A, and the magnetic field in hole B monitored with the pick-up coil of a conventional niobium flux transformer coupled to radiofrequency SQUID. The flux sensitivity was calibrated with the UPd_2Al_3 point raised, and the point was then lowered until the desired contact had been obtained. The current-voltage characteristics were monitored by a conventional four-point method using leads attached directly to the UPd_2Al_3 and niobium specimens with silver paint. The entire system was submerged in ^4He in a glass dewar. All leads were filtered against radiofrequency interference and the experiment was shielded with superconducting lead foil and mu-metal.

Figure 2 shows the flux detected at three temperatures as the flux applied by the solenoid was swept slowly up and down. It shows conventional hysteresis with steps at flux-quantum intervals. As the flux builds up in the solenoid, circulating currents around the hole containing the solenoid also build up in an attempt to keep the flux in that hole zero. When the critical current of the weak link is exceeded, a quantum of flux moves across into the second hole and is detected by the SQUID. On reducing the flux in the solenoid the reverse process occurs; the hysteresis arises from the losses in the jumping process. When the flux sweep was stopped between jumps, the trapped flux remained constant, thus demonstrating the presence of circulating currents. The slight slope in the upper and lower branches is due partly to the mutual inductance between the two holes. We observe that the critical current of the junctions is reduced as the temperature is raised. In the results shown, the modulation disappears above 1.3 K. We have studied a range of links of different strengths and have observed modulation of the detected flux up to a temperature of 1.8 K, consistent with the measured T_c for this sample (~ 2 K).

To demonstrate the quantization of flux more accurately, we stopped the flux sweep at point X. Operation of a hair drier at ~ 10 -s intervals introduced electromagnetic interference which stimulated transitions between fluxoid states as shown in Fig. 3. From the previous calibration of the SQUID sensitivity, we determined that these were of amplitude $(0.99 \pm 0.03)h/2e$.

Early measurements on this system were made by rotating the heavy-fermion screw into the slot in the niobium, to maximize the probability of forming the superconducting weak links. There is a small but worrying possibility that this could result in a thin track of niobium being deposited on the heavy-fermion point and thus forming a niobium link across the slot. To eliminate this possibility, we modified the apparatus to enable the point to be lowered into the slot without significant rotation, and quantization was confirmed. These measurements provide unambiguous confirmation of superconductivity in UPd_2Al_3 and are also, we believe, the first measurements of this kind involving a heavy-fermion system. Earlier measurements on $\text{UBe}_{13}/\text{Ta}$

FIG. 2 Flux detected by SQUID against flux applied through long solenoid. The outer loop is for $T = 1.22$ K and the middle loop is for $T = 1.25$ K. The central staircase is for $T = 1.26$ K. The steps in both axis are at intervals of one flux quantum.

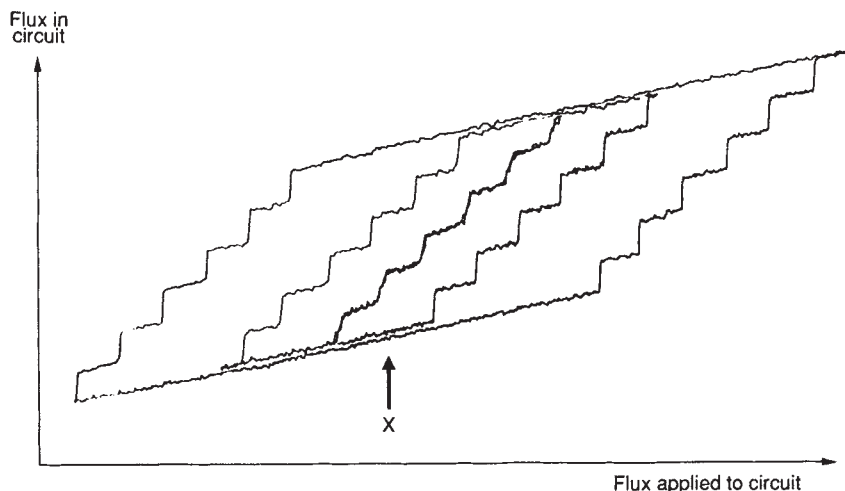
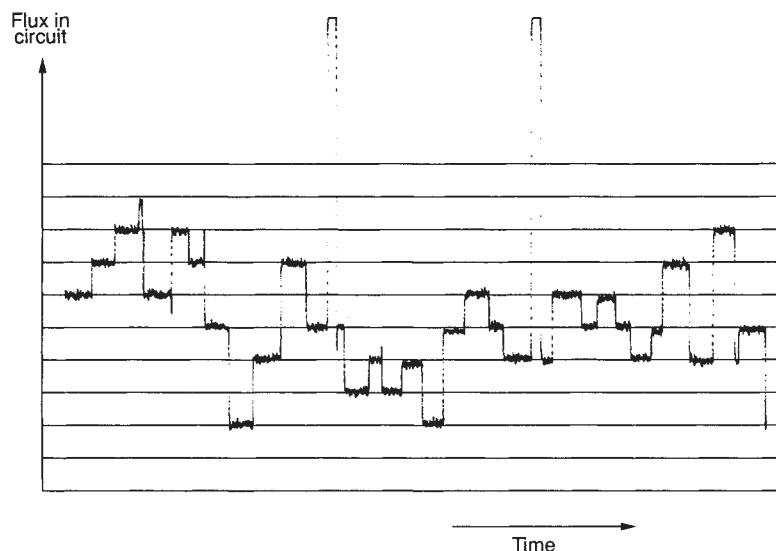


FIG. 3 Flux jumps stimulated by deliberately applied electromagnetic interference when the applied flux was at point X on Fig. 2. The separation between the grid lines is $0.99h/2e$. The two large jumps are due to unlocking of the measuring SQUID.



junctions¹¹ have shown both microwave steps and a finite resistance ($\sim 0.2 \Omega$) in series with the junction. The resistance was believed to occur between a proximity-induced *s*-state region in the UPd_2Al_3 and its supposed bulk *p*-state superconductivity. This interpretation was challenged¹² but it is worth noting that measurements of the kind reported here place a much more severe limitation on any possible series resistance associated with the interface between two wavefunctions, having the potential for different symmetry ($< 10^{-11} \Omega$ at $\sim 10 \mu\text{A}$ in our experiment), than do direct current-voltage measurements. Neither the present experiment nor that of ref. 11 constitutes a measurement of the size of the flux quantum in the heavy-fermion material itself.

Typical current-voltage curves are shown in Fig. 4. These are reversible on cycling the temperature. On some occasions we have seen Josephson-like characteristics which do not disappear

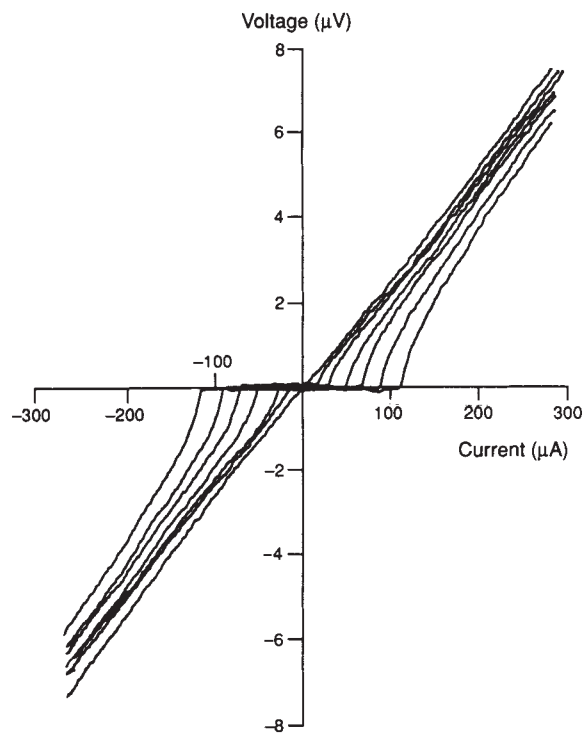


FIG. 4 Voltage-current characteristics of a weak link formed between the heavy-fermion and niobium specimens. These curves were measured at temperatures of 1.31, 1.36, 1.43, 1.50, 1.58, 1.68 and 1.8 K. $R_n \approx 25 \text{ m}\Omega$. It is not possible, in this experiment, to distinguish between $\text{UPd}_2\text{Al}_3/\text{Nb}$ and $\text{UPd}_2\text{Al}_3/\text{UPd}_2\text{Al}_3$ junctions.

above 2 K and which are probably associated with niobium-niobium contacts. We should emphasize that we cannot distinguish between $\text{UPd}_2\text{Al}_3/\text{Nb}$ and $\text{UPd}_2\text{Al}_3/\text{UPd}_2\text{Al}_3$ junctions (due to microcracks) in this experiment, but that this is unimportant for the quantization measurement. We also note that the quantization measurement is dominated by the weakest of at least two weak links in series, whereas the current-voltage characteristics are dominated by the strongest of at least two junctions in parallel. Comparison of the data from the two experiments, therefore, requires some caution. These junctions display very low $I_c R_n$ products. If we were to assume a BCS gap $\Delta = 1.76 k_B T_c$, then from the Ambegaokar-Baratoff theory for the ideal tunnelling current in BCS superconductors¹³, we would expect an $I_c R_n$ product $\sim 1 \text{ mV}$ at $T=0$. We have seen no values higher than $5 \mu\text{V}$. By contrast, 80% of the upper theoretical limit has been observed for $\text{CeCu}_{2.2}\text{Si}_2/\text{Al}$ junctions¹⁰. There are a number of reasons for low $I_c R_n$ products, even between identical conventional superconductors, so we would be reluctant to place too much weight on this result. We cannot, however, preclude the possibility that it may result from an unusually small tunnelling matrix element between the superconducting states in the two materials. Further studies of the current-voltage characteristics are in progress.

We have also attempted to establish quantization in a composite ring of UPd_2Al_3 and the high-temperature superconductor $\text{YB}_2\text{C}_3\text{O}_7$, but have so far been unsuccessful in obtaining either persistent currents or Josephson-like characteristics. This may simply result from a difficulty in producing point contact between $\text{YB}_2\text{C}_3\text{O}_7$ and other brittle materials, so we have tried using gold leaf and also sputtered gold as intermediaries in an attempt to produce proximity-coupled junctions, but without success. Further efforts are in hand, but again we cannot preclude the possibility that this negative result is connected with some difference in the nature of the wavefunctions in the two materials. $\square$

Received 5 March; accepted 3 April 1992.

1. Steglich, F. *et al.* *Phys. Rev. Lett.* **43**, 1892-1896 (1979).
2. Flouquet, J. *et al.* *Physica C* **185-189**, 372-378 (1991).
3. Geibel, C. *et al.* *Physica C* **185-189**, 2651-2652 (1991).
4. Geibel, C. *et al.* *Z. Phys.* **B84**, 1-3 (1991).
5. Gough, C. E. *et al.* *Nature* **326**, 855 (1987).
6. Keene, M. N., Jackson, T. J. & Gough, C. E. *Nature* **340**, 210-211 (1989).
7. Rice, T. M. *Jap. J. Appl. Phys.* **26** (Suppl. 26-3) 1865-1874 (1987).
8. Chaudhari, P., Dimos, D. & Mannhart, J. *IBM J. Res. Dev.* **33**, 299-305 (1989).
9. Wilczek, F. *Int. J. Mod. Phys. B* **5**, 1273-1312 (1991).
10. Steglich, F. *et al.* *J. appl. Phys.* **57**, 3054-3059 (1985).
11. Han, S., Ng, K. W. & Wolf, E. L. *Phys. Rev. Lett.* **57**, 238-241 (1986).
12. Kadlin, A. M. & Goldman, A. M. *Phys. Rev. Lett.* **58**, 2275 (1987).
13. Ambegaokar, V. & Baratoff, A. *Phys. Rev. Lett.* **10**, 486-489 (1963).

ACKNOWLEDGEMENTS. We thank M. Gunn, M. Long and C. Gough for useful discussions and E. Forgan for helping to set up the collaboration.

Effect of deep convection on the regulation of tropical sea surface temperature

John M. Wallace

Department of Atmospheric Sciences, University of Washington, Seattle, Washington 98195, USA

THE distribution of tropical sea surface temperatures (SSTs) is negatively skewed: a 'warm pool' with SSTs greater than 300.5 K covers roughly half the tropical oceans (15° N to 15° S), whereas SSTs as low as 293 K are observed in regions of equatorial and coastal upwelling and persistent stratus cloud cover¹. SSTs are thus within 2–3 K of the highest values over a large area of the tropical ocean, leading some authors to suggest that some physical process may act to limit SSTs to below about 303 K. Ramanathan and Collins² have proposed a 'thermostat' mechanism in which ocean warming produces enhanced deep convection, leading to the formation of extensive cirrus cloud canopies which shield the troposphere and ocean surface from incoming solar radiation. Here I suggest that a mechanism of this sort may not be required to explain the SST distribution. I argue that large-scale dynamical processes will act to maintain uniform tropical tropospheric temperatures to within about 2 K, and that, in the absence of horizontal temperature contrasts in the atmosphere, a negatively skewed SST frequency distribution is bound to develop through equilibration between the atmosphere and spatially varying SSTs. In addition, although cirrus clouds reduce the solar insolation at the Earth's surface in regions of deep convection, they would not necessarily prevent SSTs from exceeding 305 K in the face of extensive greenhouse warming.

The nonlinearity inherent in the Clausius–Clapeyron equation has sometimes been invoked to explain the negatively skewed SST distribution³. The fact that the saturation vapour pressure over water increases with SST suggests that latent heat fluxes may become increasingly efficient at cooling the ocean surface as SST increases. This effect may contribute to the observed skewness of SST but it seems unlikely that this rather weak nonlinearity, by itself, could account for the narrowness of the mode in the observed frequency distribution.

As a consequence of dynamical constraints, pressure surfaces in the tropical troposphere are remarkably flat (the zonally symmetric case has been discussed in refs 4–6). For example, Fig. 1 shows the April climatological mean geopotential height field on the 200 hPa surface, which is situated in the upper

troposphere at the level of strongest horizontal gradients. The range between the highest and lowest values in the tropics is not more than two contour intervals (80 m). Maps for individual years/months at levels ranging from the surface to the tropopause are similar in this respect. As the vertical scale of tropical circulation systems is comparable to the depth of the troposphere it follows from the hypsometric equation that the mean (vertically averaged) temperature of the tropical troposphere must be uniform to within ~2 K, as illustrated by the observational example shown in Fig. 2. Tropical temperatures tend to be more horizontally uniform in the free atmosphere (above 850 hPa) than in the planetary boundary layer, which feels the larger spatial variability of the underlying SST field.

From the standpoint of the coupled climate system, the tropical atmosphere behaves as it diffused heat virtually instantaneously, so as to eliminate horizontal temperature gradients. Heat is horizontally redistributed, not by turbulent eddies, but by thermally direct circulations, with rising of warmer air, sinking of cooler air, and a down-gradient horizontal transport of moist static energy.

Now consider a hypothetical tropical ocean domain whose average SST is in radiative convective equilibrium⁷ with an overlying atmosphere whose temperature and radiative properties are horizontally uniform. But in contrast to observations, let us suppose that SST starts out as being normally distributed, so that regional 'hot patches' are just as far above the spatial mean as the patches of cold water currently observed in the eastern Atlantic and Pacific are below it. The departures of SST from the domain average value (hereafter denoted by SST*) would perturb the distribution of energy fluxes at the air–sea interface which, in turn, would modify the distribution of deep convection and induce large-scale circulations within the atmosphere. The fluxes at the air–sea interface would also modify the underlying distribution of SST*. I would argue that the resulting feedbacks are bound to induce negative skewness into the frequency distribution of SST*.

In this scenario, hot patches in the SST* distribution would be subject to enhanced fluxes of latent and sensible heat at the air–sea interface, which would locally cool the ocean and heat the atmospheric planetary boundary layer, destabilizing the overlying free troposphere. These regions would thus become preferred sites for deep convection, which would have otherwise been randomly distributed over the domain. The excess energy taken from the ocean surface would not remain within the planetary boundary layer: it would be immediately carried upwards by the convection. It would heat the troposphere not just over the local hot patches, but over the entire domain. Hence, the air–sea fluxes would continue to be enhanced so

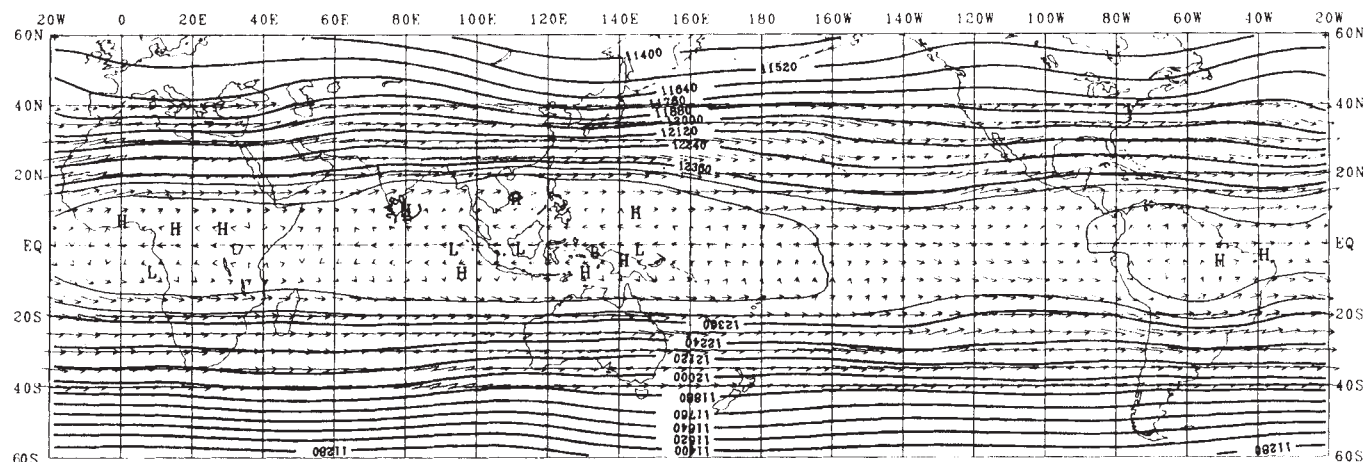


FIG. 1 Climatological mean 200 hPa geopotential height field for April. The contour interval in the tropics is 40 m. From ref. 8.

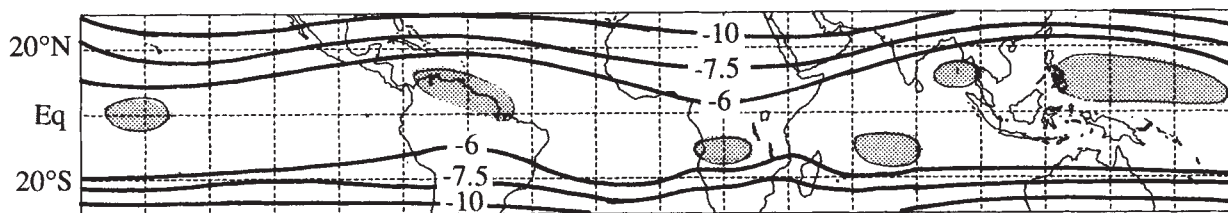


FIG. 2 Climatological mean 500 hPa temperature height field for March–May. The contour interval in the tropics is 1 °C. The contour interval is variable and values between –4 and –5 °C are shaded. Adapted from ref. 9.

long as a hot patch remained appreciably warmer than the domain-averaged radiative–convective equilibrium temperature.

In a similar manner, the air–sea fluxes would initially be suppressed in regions of low SST*, locally cooling the atmospheric planetary boundary layer and warming the ocean relative to the domain average. But as the stratification over these regions is stable, there would be nothing to prevent the atmospheric planetary boundary layer from equilibrating with the underlying SST*. Beyond this rather short equilibration time, the suppression of the air–sea fluxes over the cold patches would not be as pronounced as the enhancement of the fluxes over the hot patches. Hence, the hot patches would disappear more quickly, leaving a prominent mode in the frequency distribution of SST not far below the highest observed SST, thereby creating the impression of a homogeneous ‘warm pool’, in (or close to) radiative–convective equilibrium with the overlying, uniform atmosphere. In the presence of dynamical processes in the ocean that generate horizontal inhomogeneities in the SST distribution, this equilibration process would maintain the observed negative skewness in the frequency distribution of SST, even in the absence of nonlinearities in the Clausius–Clapeyron equation or cloud feedbacks.

Observational evidence has been presented (in ref. 2 for example) to the effect that SST and highly reflective cloud cover tend to be positively correlated over the tropical Pacific. Ramanathan and Collins² make the important point that, on average, the solar radiation incident on the ocean surface in the warm pool tends to be substantially less than that on regions of lower SST. I agree that in the absence of the spatial gradients in cloud cover the energy balance over the warm pool would be much different from what we observe today. But I doubt very much that SST would get out of control, as implied by their use of terms such as ‘super greenhouse’ and ‘thermostat’. The energy balance considerations, on which their analysis is based, are insufficient for inferring causality or making predictions as to how the system would behave in the absence of some process. Without a more comprehensive analysis based on a dynamical model of the coupled atmosphere–ocean system, there is no way of knowing to what extent artificially erasing the contrast in insolation between the areas of deep convection over the warm pool and the equatorial dry zone to the east would be compensated by the inevitable adjustments in the atmospheric circulation and the associated wind-driven ocean circulation and air–sea fluxes. Given the horizontal homogeneity of tropospheric temperatures, it seems unlikely that SST over any part of the warm pool could be markedly warmer than it is today unless the domain average radiative balance were altered.

The report of Ramanathan and Collins² invites the further inference that in the event of changes in the global energy balance (for example, greenhouse warming), the same ‘regulatory effect’ of cirrus clouds may limit SST to less than 305 K. In this respect, it should be emphasized that their analysis does not address the difficult question of whether the fractional area of the tropics covered by highly reflective clouds would change in response to a rise in the radiative–convective equilibrium temperature of the tropics. Unless cloudiness were to increase over the tropics as a whole, it is not clear how the radiative–

convective equilibrium temperature of the tropical oceans could be held in check by their negative feedback mechanism. □

Received 13 September 1991; accepted 6 April 1992.

1. Sadler, J. C., Lander, M. A., Hori, A. M. & Oda, L. K. *Tropical Mean Climatic Atlas Vols 1 and 2* (HMET 87-01 and 87-02 (Department of Meteorology, Univ. of Hawaii, 1987).
2. Ramanathan, V. & Collins, W. *Nature* **351**, 27–32 (1991).
3. Newell, R. E., Navato, A. R. & Hsiung, J. *Pure appl. Geophys.* **116**, 351–371 (1978).
4. Schneider, E. K. *J. Atmos. Sci.* **34**, 280–297 (1977).
5. Held, I. M. & Hou, A. Y. *J. Atmos. Sci.* **37**, 515–533 (1978).
6. Lindzen, R. S. & Hou, A. Y. *J. Atmos. Sci.* **45**, 2416–2427 (1988).
7. Sarachik, E. S. *Dyn. Atm. Oceans* **2**, 455–469 (1978).
8. Japan Meteorological Agency *Monthly Report on the Climate System Tech. note 35* (Tokyo, 1992).
9. Newell, R. E., Kidson, J. W., Vincent, D. G. & Boer, G. J. *The General Circulation of the Tropical Atmosphere* Vol. 1 (MIT Press, Cambridge, Massachusetts, 1972).

ACKNOWLEDGEMENTS. This note arose out of stimulating conversations with E. Sarachik and D. L. Hartmann. This work was supported by the US NSF.

Importance of gravitational spreading in the tectonic and volcanic evolution of Mount Etna

A. Borgia*, L. Ferrari† & G. Pasquare†

* Department of Earth Sciences, The Open University, Milton Keynes MK7 6AA, UK

† Dipartimento Scienze della Terra, Università degli Studi di Milano, via Mangiagalli 34, 20133 Milano, Italy

THE interaction between gravity and thermal effects largely determines the structural and magmatic evolution of volcanic constructs at scales spanning several orders of magnitude, from small cones to the oceanic crust^{1–3}. Although gravitational spreading is a direct consequence of this interaction and a fundamental process in volcano growth, it is rarely recognized as such. Here we describe a striking example of this process, at Etna volcano, in Italy. The volcano and its clay-rich substratum are slowly spreading towards the east and south, driven by gravity. Spreading produces extensional structures in the summit region and compressional structures at the base of the volcano. Eastward movement of the volcanic edifice over a stationary magma supply may be the cause of an apparent westward migration of volcanic activity. As gravitational spreading seems to control the location and magnitude of shallow seismicity and flank eruptions, an appreciation of its effects could become an essential element of future volcanic hazard evaluation. Our model proposed here for Etna may also be relevant to a reinterpretation of the geological history of a number of other volcanoes.

Gravitational spreading of volcanic edifices produces two main structural features by which the process may be identified. These are, first, summit grabens and rifts (the product of extension), and second, basal thrust faults and folds (the products of compression). The association of these features with gravitational spreading has been recognized for only a few terrestrial and martian volcanoes^{4–8}. We propose that at Mount

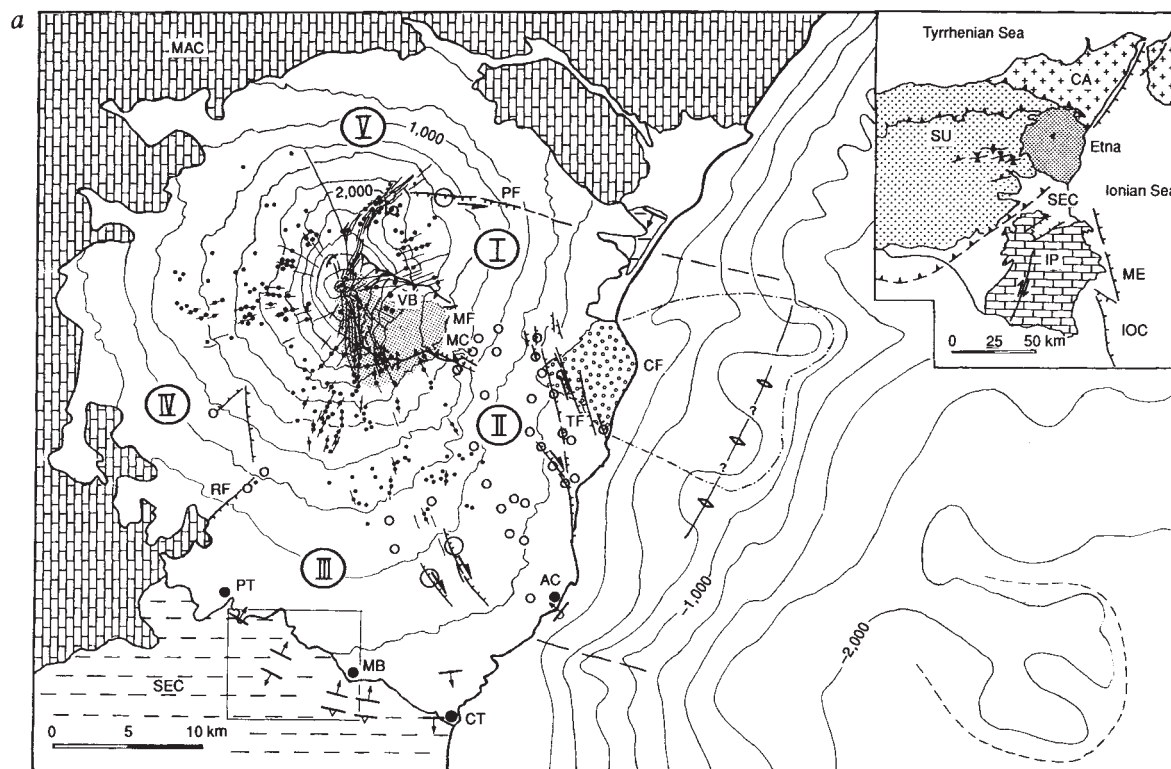
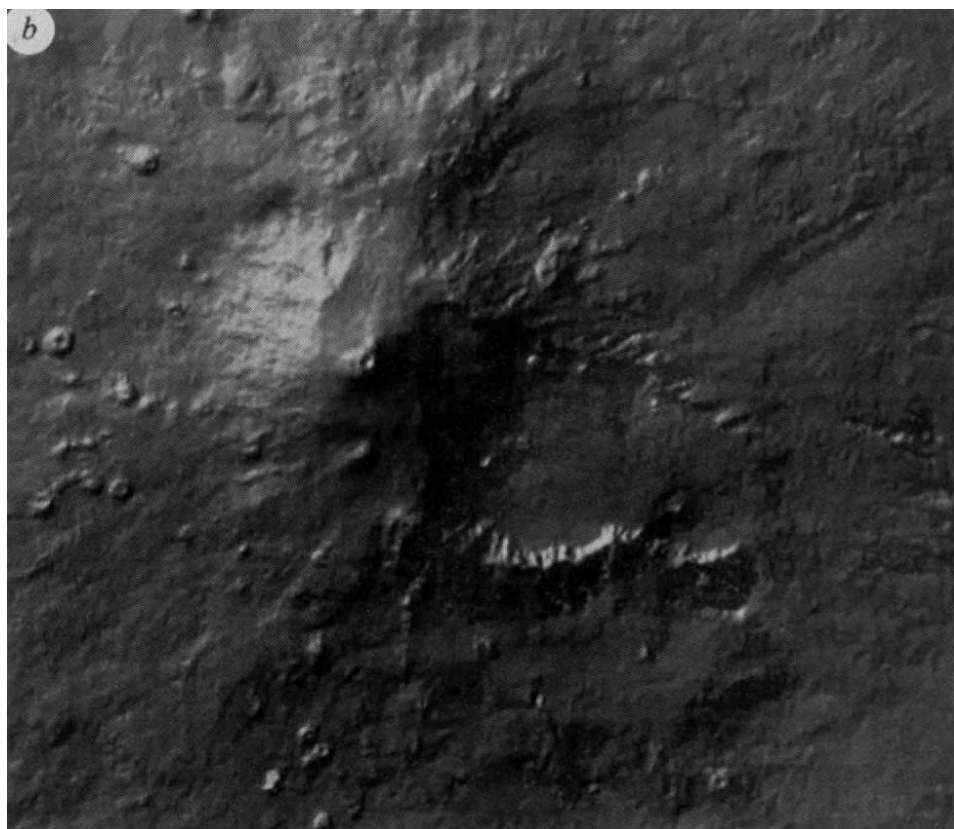


FIG. 1 *a*, Simplified geological map of Etna Volcano. Etna is located on the southern border of the Maghrebian-Apenninic Chain (MAC), composed of the Calabrian Arc (CA) and the Sicilide Units (SU). South of the chain, the carbonate and volcanic rocks of the Iblean plateau (IP) crop out. In between, the Catania basin is filled with sand and clay units, known in the upper part of the sequence as the sub-etnean clays (SEC). The Iblean plateau terminates to the west of Etna along a transpressional fault zone²⁷; to the east, it is bounded by the extensional zone of the Malta Escarpment (ME) which borders the Ionian ocean crust (IOC)²⁸. Etna^{26,29} has four summit craters and three rift zones. The northeast and west rifts seem to abut against the Maghrebian-Apenninic Chain where the Pernicana fault (PF)³⁰ and the Ragalna fault system (RF)²⁹ begin. These faults continue to the east and south, respectively, and may follow the boundary between the chain and the sub-etnean clays^{9,29}. Instead, the south rift terminates on the lower slopes of the volcano in an area with widespread eruptive vents²⁹ and normal right-lateral faults²¹. The highest concentration of superficial earthquakes^{31,32} and the Timpe normal faults (TF) are found just northeast of this area. In the Valle del Bove (VB) topographic depression, most normal faults, dykes and eruptive fractures have a dominant trend of 60° on the northern wall of the Valle and a trend of 140° on the southern wall¹⁰; a few also parallel the trend of the northern wall²⁶. AC, Aci Castello; CT, Catania; MB, Misterbianco; MC, Monte Calanna; MF, Monte Fontane; PT, Paterno. Ticked lines indicate active faults, with sense of displacement where known. Thin lines indicate eruptive fractures; dots show eruptive vents. Dotted pattern in Valle del Bove indicates extent of plutonic complex at 5 km depth b.s.l.¹⁸. Region with pattern of circles shows outcrop of Chiancone fanglomerate (CF); dot-dashed line shows inferred seaward extent of Chiancone fanglomerate. Dashed line



shows possible submarine landslide. Large circles indicate macroseismic epicentres³². North is up. Box indicates area of Fig. 2. *b*, Digital elevation model of the summit area of Etna Volcano. Note the extensional fault systems striking north-south and west-northwest-east-southeast, which are thought to originate from gravitational spreading of the volcanic edifice and to control the erosion of Valle del Bove, the large depression southeast of the summit (VB in *a*). North is up. Images were processed by G. Macedonio, M. T. Pareschi and P. Armienti at Università degli Studi di Pisa.

Etna, the Valle del Bove and the basal anticlines indicate summit extension and basal compression, respectively, produced by the gravitational spreading process.

The Valle del Bove is a 7-km-long topographic depression which strikes east-southeast from the summit (Fig. 1a and b). The northern and western walls of Valle del Bove are fairly straight, but the southern wall is U-shaped, making the Valle twice as wide towards its head as towards its exit. The western and southern walls are rimmed by normal faults^{9,10}. Caldera eruption, landsliding and glacial erosion have been suggested as causes for the formation of this depression^{9,11,12}; whatever the mechanism, the Chiancone coastal flanglomerate (Fig. 1a) is thought to have originated from massive erosion during the formation of the Valle^{9,11}.

The calculated minimum east-west extension of the summit area due to dyke intrusion over the past 34 kyr is ~500 m, whereas north-south extension amounts to 400 m (ref. 13). We therefore deduce an average minimum rate of extension of 1.3 cm yr^{-1} . This value compares well with that of 1 cm yr^{-1} obtained from preliminary Global Positioning System measurements¹⁴. Whereas extension characterizes the more recent history of the upper slopes of Etna, the older sequences at Monte Calanna (Fig. 1a) provide evidence of an older compressional phase¹⁵. A gravity high, located in the southern central part of Valle del Bove^{16,17}, corresponds to a high in seismic velocity¹⁸. These anomalies are interpreted as a plutonic complex located from 2 to 6 km below the surface.

A variety of tectonic regimes occurs along the southern base of Etna^{19,20}. A broad anticline trending east-west (Fig. 1) folds the sub-etnean clays, indicating a north-south compression. The clay-rich nucleus of this anticline, however, has undergone pervasive ductile and brittle deformation suggestive of eastward tectonic transport. Just south of the anticline, active extension in the direction east-southeast to west-northwest is found¹⁹. The anticline was interpreted²⁰ as a product of the last north-south compressional phase at the front of the Maghrebian-Appennine accretionary prism, whereas the eastward tectonic transport was thought to be due to lateral gravitational gliding over a slope generated by side ramps of the prism or by the Malta escarpment.

Sub-etnean clays and basal tholeiitic lavas along the eastern and southeastern coast of Etna are uplifted more than 100 m and substantially tilted (Fig. 1a; ref. 9). Outcrops in Catania show southward-dipping strata, and at Aci Castello pillow lavas embedded in the clays are overturned to the southeast. This deformation has been interpreted as the result of regional uplift of the etnean domain associated with down-faulting of crustal blocks⁹. The blocks were thought to slide eastwards on a decollement which was supposed to extend for many tens of kilometres along the base of the crust (20 km below the sea level; ref. 9). Dyke intrusion along the rifts, displacement along the faults and seismicity are consistent with seaward movement of the eastern flank of Etna^{9,21,22} in a way comparable to that of the southern flank of Kilauea⁶.

In addition to the outcrops above, we surveyed other areas which are instrumental for understanding the genesis of the Valle del Bove and of the basal anticlines. In the Valle del Bove at Monte Fontane (Fig. 1a) the 80-kyr-old²³ volcanic sequence unconformably overlies an older, 400-m-thick sequence dipping from 95° to 75° . The tilted sequence suggests a shortening and an uplift both of ~100 m. The sequences are cut by normal faults trending 30° with downthrows towards the west, and trending 90° with downthrows toward the Valle del Bove. This tectonic setting is similar to that observed in the Monte Calanna area where an older compressional phase is followed by younger extension. The existence of compressive tectonics on the intermediate slopes of a volcano is not surprising, and although rarely recognized on terrestrial volcanoes, it is predicted by theory^{8,24,25} and has been suggested for Olympus Mons on Mars^{24,25}.

Although no landslide deposits originating from sector collapses from the Valle del Bove crop out, the reworked material of the Chiancone flanglomerate, which seems to extend up to 10 km off-shore (Fig. 1a), has an estimated volume comparable to that of the Valle del Bove. In addition, the submarine topographic high (possibly a landslide) marked by the contour at 2,000 m below sea level (b.s.l.) shows no direct connection with the Valle del Bove, but seems to originate from a headscarp below sea level. These observations, in addition to the presence of boundary normal faults parallel to the walls of Valle del Bove (Fig. 1), and the fact that the Valle becomes narrower toward its exit, are evidence against genesis of the Valle by sector collapse. Instead, they favour an origin by massive erosion (including minor landsliding) which is structurally controlled.

Geophysical and structural studies of the sub-etnean clays south and east of Etna (Figs 1 and 2) show that the basal anticline dies out northwestward towards the Ragalna fault, although it seems to extend eastwards below Catania, with a southward vergence, and northeastward to Aci Castello, with a southeast vergence. The southern limb of this anticline forms a kink fold that makes the strata vertical. Seismic and borehole data indicate a structural uplift of ~500 m (Fig. 2). In the cross-section of Fig. 2b, the structure is interpreted as a fault-propagation fold generated by north-south compression. A thrust dipping at $0^\circ/5^\circ$ and located at 3 km depth steepens upward to 20° below Misterbianco, producing the ~500 m of shortening and uplift observed at the surface. The calculated thrust depth corresponds to the top of the Iblean plateau, which can be extrapolated below Etna to the base of the plutonic complex (Figs 2b and 3b). The anticline is younger than 300 kyr old (as etnean lavas of this age are involved in the deformation), and it has been active until recently, judging from the observed syn-sedimentary tectonic deformation. The minimum rate of thrusting is therefore calculated as 0.2 cm yr^{-1} . Given that the calculated thrusting is only slightly larger than the summit north-south extension due to dyke intrusion (~400 m) and that the processes should have overlapped in time, we suggest that a relevant part of the observed thrusting at the base of the volcano is compensated by summit extension associated with southward spreading of the flank of Etna. The superficial stress associated with this north-south slumping of the flank could mask the actual deep regional tectonic stress, which is extensive¹⁹.

A ridge (from 700 to 1,900 m b.s.l.) extends off-shore, along the base of Etna, northwestward from Aci Castello to the latitude of the Pernicana fault (Fig. 1a). Its morphology is similar to those at the base of Poás and Kilauea volcanoes⁷. Given that the flank of Etna is moving outward like those of these volcanoes^{21,22,26}, we interpret the step at Etna by analogy as the morphological expression of an active compressional structure, the off-shore continuation of the anticlines found south of Etna. If this is valid, the height of the scarp and the wavelength of the bench give a minimum of 1.2 km of thrusting at the depth of the Ionian crust (~4 km b.s.l.). The lavas of Aci Castello²³ limit the maximum age for the beginning of deformation and the minimum rate of thrusting to 500 kyr and 0.2 cm yr^{-1} , respectively. As for the southern flank, compression and shortening along the base is accompanied by a minimum of 500 m extension in the summit area¹³.

We propose that Etna is slowly spreading radially to the east and south, although it is buttressed by the Maghrebian-Appennine Chain to the north and west (Figs 1 and 3). The spreading produces extension at the summit and compression at the base, and seems to vary between adjacent sectors (blocks), being greater towards the east. It seems to be decoupled by thrust faults (a decollement) from the Iblean plateau and the Ionian ocean crust at a depth of ~5 km b.s.l., with an estimated minimum thrusting of 1.2 km eastward and 0.5 km southward. The thrusts ramp upward along the southern and eastern base generating broad anticlines. The spreading region seems

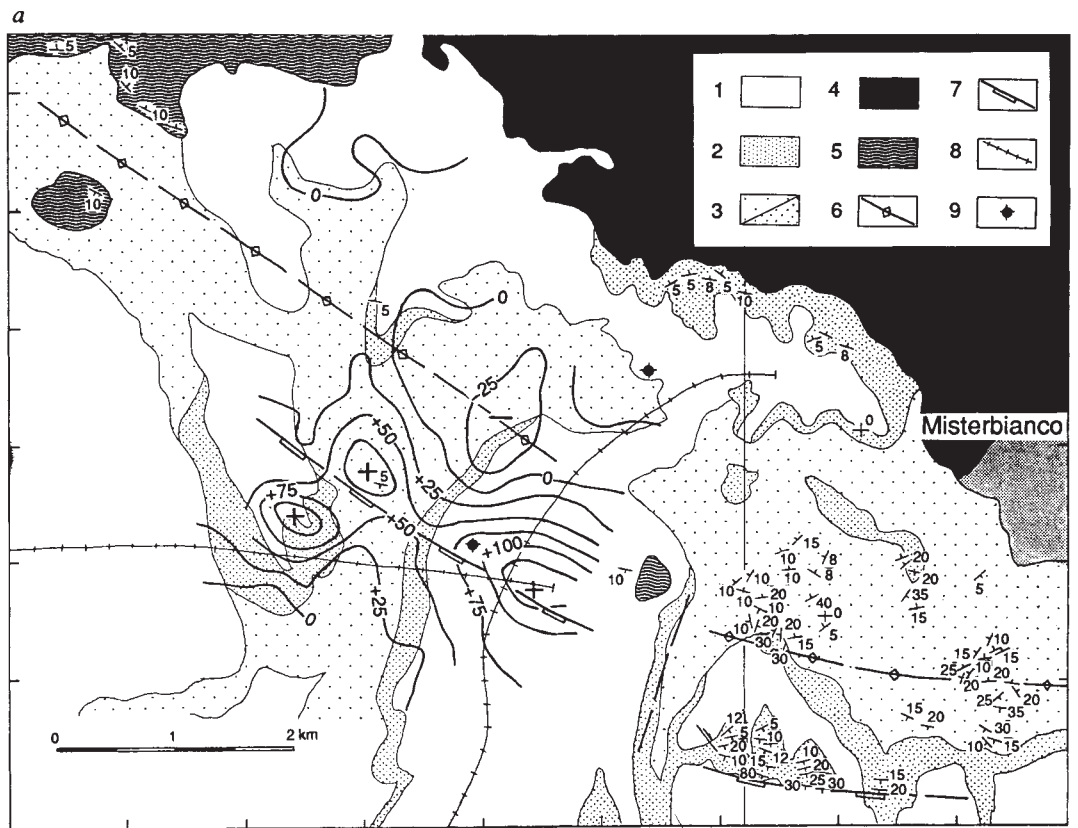
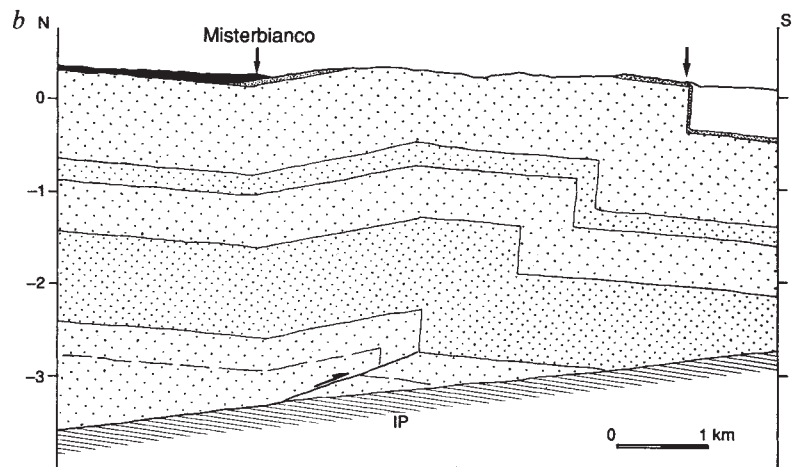


FIG. 2a, Geological map of the anticline south of Etna with structural and geophysical data. Location is indicated in Fig. 1. North is up. Key: 1, recent sedimentary deposits; 2, sands and conglomerates; 3, sub-ethnean clay-rich (light stipples) and sand-rich (dense stipples) units; 4, recent etnean lavas; 5, basal tholeiites; 6, suggested culmination of the anticline; 7, inferred hinge of the kink fold; 8, trace of seismic profile (AGIP); 9, borehole (AGIP). Thin line is trace of cross-section in *b*. The irregular attitudes of the clay beds are thought to represent local deformation induced by the proximity of the topographic surface and are to a first approximation neglected in the cross-section (*b*). Contour lines (at intervals of 25 nanotesla) of the total magnetic field are from a preliminary high-resolution (more than 1,000 data points) magnetic survey of the anticline, filtered for the higher frequencies. The anomaly is produced by the basal tholeiites encountered in the southern borehole, which are seen displaced in the seismic profile. The main axis of this anomaly is parallel to the hinge of the anticline. *b*, Balanced cross-section³³ of the anticline based on conservation of bed thickness and length. The actual structure is probably more complex than the one proposed but cannot be better constrained by the available data. Trace of section is indicated by the thin line in *a*. Patterns are as in *a*. IP, Iblean plateau. Stratigraphy is based on AGIP borehole data. Arrow on the right locates the outcrop shown in *c*. *c*, Outcrop of low-angle thrusting and kink-folding in the sub-ethnean clays (SEC) located by the arrow in *b*; note the wedge squeezed out of the fold hinge.



confined laterally by the Pernicana fault to the northwest, by the north and west-southwest rifts through the summit, and by the Ragalna fault system to the southwest. Spreading probably started 300 kyr ago with minimum rates of 0.2 cm yr^{-1} and is still active. In fact, preliminary global positioning system measurements¹⁴ and the calculated summit extension caused by dyke intrusion¹³ suggest a rate that is at least an order of magnitude larger. While sectors I and III slide east and southward, respectively, sector II is governed by motion and tectonics transitional between the two. The motion is generally towards the southeast and the tectonic is dominated by northwest-southeast compression and northeast-southwest extension. This stress regime could be responsible for the larger number of earthquakes and eruptive vents in this sector (Fig. 1a).

Along the western and northern sectors of Etna the magnitude of the deformation¹⁹ is not comparable to that observed on the eastern and southern flanks. We believe that this difference in tectonic regime is a consequence of the different topography and mechanical properties of the substratum (Fig. 3). The northern and western sectors overlay and abut against the topographically high, fairly strong sediments of the Maghrebian-Apenninic Chain; in contrast, the other sectors stand over the topographi-

cally low, ductile clay-rich units of the sub-Etnean clays which deform under the load of the volcano. □

Received 30 September 1991; accepted 24 March 1992.

- Shaw, H. *U.S. Geol. Surv. Prof. Pap.* 1350, 1357-1375 (1987).
- Shaw, H. & Chouet, B. *J. geophys. Res.* **96**, 10191-10208 (1991).
- Borgia, A. & Treves, B. *J. geol. Soc. Lond.* **60**, 277-300 (1992).
- van Bemmelen, R. W. *The Geology of Indonesia* (Government Printing Office, La Hague, 1949).
- Suzuki, T. *Bull. Volcanol. Soc. Japan* **13**, 95-108 (1968).
- Swanson, D. A., Duffield, W. A. & Fiske, R. S. *U.S. Geol. Surv. Prof. Pap.* 963 (1976).
- Borgia, A., Burr, J., Montero, W., Morales, L. D. & Alvarado, G. E. *J. geophys. Res.* **95**, 14357-14382 (1990).
- Borgia, A. & Melosh, J. *Proc. Int. Conf. Active Volcanoes and Risks Mitigation*, 3 (Osservatorio Vesuviano, Napoli, 1991).
- Kieffer, G. thesis, Clermont-Ferrand Univ. (1985).
- Calvari, S., Groppe, G. & Pasquaré, G. *Acta vulcanol.* (in the press).
- Guest, J. E., Chester, D. R. & Duncan, A. M. *J. Volcan. geotherm. Res.* **21**, 1-23 (1984).
- McGuire, W., Murray, J., Pullen, A. & Saunders, S. *J. geol. Soc. Lond.* **148**, 577-583 (1991).
- Ferrari, L., Garduño, V. H. & Neri, M. *Proc. Congr. Soc. Geol. Ital. 'Leo Ogniben'*, 222 (Univ. of Catania, 1991).
- Nunnari, G. & Puglisi, G. *Acta Vulcanol.* **1**, 268-272 (1991).
- Ferrari, L. *et al. Boll. Grup. Naz. Vulcan.* 849-860 (1989).
- Neumann, R., Everaud, P. & Coron, S. *Boll. Geof. teor. appl.* **106**, 125-152 (1985).
- Loddo, M., Patella, D., Quarto, R., Ruina, G., Tramaccare, A. & Zito, G. *J. Volcanol. geotherm. Res.* **39**, 17-39 (1989).
- Hirn, A., Nercissian, A., Sapin, M., Ferrucci, F. & Wittlinger, G. *Geophys. J. Int.* **105**, 139-153 (1991).
- Bousquet, J. C., Gresta, S., Lanzafame, G. & Paquin, C. *Mem. Soc. geol. Ital.* **38**, 483-506 (1987).
- Labenne, P., Bousquet, J. C. & Lanzafame, G. *Tectonophysics* **177**, 349-366 (1990).
- Lo Giudice, E. & Rasà, R. *Tectonophysics* **202**, 257-268 (1992).
- Neri, M., Garduño, V. H., Pasquaré, G. & Rasà, R. *Acta vulcanol.* **1**, 17-24 (1991).
- Romano, R. & Gillot, P. *Proc. Hawaii Symp. on How Volcanoes Work*, 211 (Hawaiian Volcano Observatory, Hilo, Hawaii, 1987).
- Thomas, P. J., Squyres, S. W. & Carr, M. H. *J. geophys. Res.* **95**, B9 (1990).
- McGovern, P. J. & Salomon, S. C. *Lunar Planet. Inst. Tech. Rep.* 90-06, 217-218 (1990).
- Kieffer, G. *Rev. Géol. Dyn. Géogr. Phys.* **24**, 129-152 (1983).
- Bianchi, F. *et al. Mem. Soc. geol. Ital.* **38**, 429-458 (1987).
- Ben-Avraham, Z. & Grasso, M. *Ann. Tectonicae* **4**, 131-139 (1990).
- Romano, R. (ed.) *Geological Map of Mt Etna* (CNR, 1979).
- Azzaro, R., Lo Giudice, E. & Rasà, R. *Boll. Grup. Naz. Vulcan.* CNR, 22-40 (1988).
- Gresta, S. & Patané, G. *Pure appl. Geophys.* **125**, 951-970 (1987).
- Azzaro, R., Lo Giudice, E. & Rasà, R. *Boll. Grup. Naz. Vulcan.* CNR, 13-46 (1989).
- Suppe, J. *Principles of Structural Geology* (Prentice-Hall, Englewood Cliffs, 1985).

ACKNOWLEDGEMENTS. We thank G. Brown, S. Calvari, F. Ferrucci, V. H. Garduño, G. Groppe, G. Kieffer, W. McGuire, J. Murray, M. Neri, M. Pompilio and H. Rymer for discussion, and Di Cesari and AGIP for permission to study boreholes and seismic profiles. This work was supported by CNR-Gruppo Nazionale per la Vulcanologia, EPOCH, and Alenia. L.F. thanks MPI for a PhD grant.

The timing of mineral growth across a regional metamorphic sequence

Kevin W. Burton & R. Keith O'Nions

Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

IF regional metamorphism occurs in response to crustal thickening, then low-grade rocks are likely to record earlier thermal conditions than will high-grade rocks in the same exhumed terrain¹⁻⁴. At Sulitjelma, North Norway, regional metamorphism during the Caledonian orogeny created a type sequence of progressive metamorphic zones⁵⁻⁷. Garnet and coexisting minerals from the high-grade part of this terrain yield precise and concordant Sm-Nd, U-Pb and Rb-Sr ages which relate directly to the pressure-temperature conditions of mineral growth⁸. Here we present pressure-temperature-time (P - T - t) information for lower-grade rocks occurring at the garnet isograd. Taken together, these data provide a precise estimate of the timing of mineral equilibration across a prograde metamorphic terrain and indicate that 'peak' metamorphic conditions at the garnet isograd (at $458 \pm 20^\circ \text{C}$) were attained $8.3 \pm 3.0 \text{ Myr}$ (Δt , Sm-Nd) or $8.7 \pm 3.2 \text{ Myr}$ (Δt , U-Pb) before the 'peak' in the high-grade sample (at $544 \pm 16^\circ \text{C}$). These results clearly illustrate that the exhumed ' P - T array' recorded in such terrains cannot be used to deduce any single geotherm that existed during metamorphism.

The Sulitjelma 'ophiolite', North Norway (67°N , 16°E), intrudes a supracrustal sequence, the Skaiti Supergroup, and is

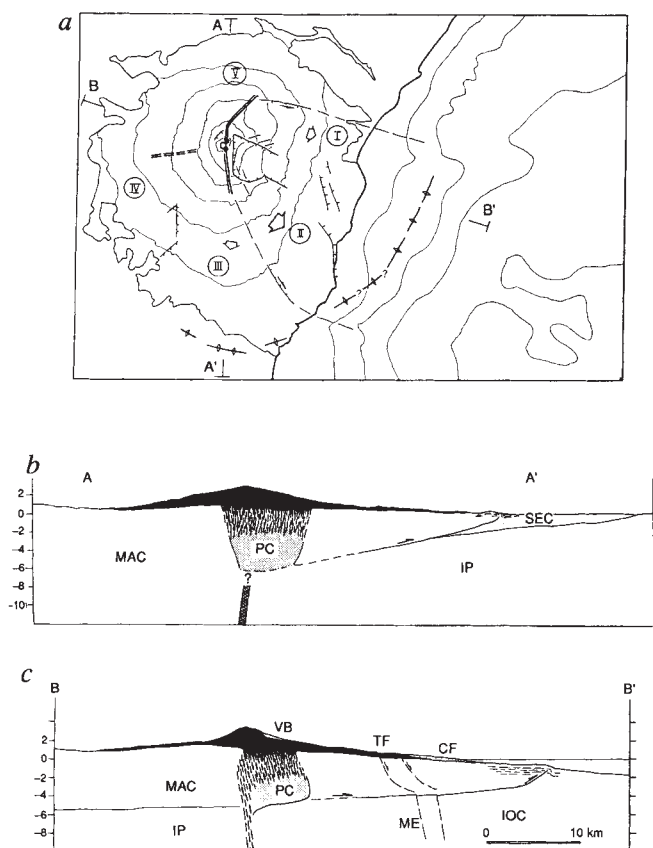


FIG. 3a, Suggested structural model for Etna; roman numerals indicate sectors (possibly blocks) with different tectonic regimes described in the text: I, eastern; II, southeastern; III, southern; IV, western; V, northern. Arrows show inferred relative displacement of sectors. Double thick lines indicate major rift zones. North is up. b, c, Schematic interpretative cross-sections of Etna volcano. Section traces are indicated in a. Note the plutonic complex (PC) passing upwards to the dyke complex and the Timpe faults thought to result from differential compaction of the flank of the volcano as it is thrust over the Malta escarpment. Patterns and acronyms as in Fig. 1; cross-hatched region indicates transpressional fault zone separating the Iblean plateau from the Maghrebian-Apennine chain²⁷. Depth of Iblean plateau inferred from ref. 27.

itself overlain by a sedimentary cover, the Furulund and Sjonsta Groups (Fig. 1). All these rocks underwent metamorphism and deformation during the Caledonian orogeny and preserve a type sequence of regional metamorphic zones which range from greenschist to amphibolite-facies⁵⁻⁷. At these metamorphic grades, garnet can provide quantitative estimates of the P - T conditions of mineral equilibration^{9,10}. Garnet also has high Sm/Nd and U/Pb ratios (and correspondingly low Rb/Sr ratios) relative to coexisting phases^{8,11-15}, and can yield precise chronological information. The diffusion rates of trace elements in garnet¹¹⁻¹⁷ appear to be at least as low as those of major cations^{18,19}. As a consequence, if the garnet composition is not modified by diffusion after its growth, then both the P - T estimates and the age information are directly related to mineral crystallization. Garnet and coexisting minerals are capable of yielding ages with a typical precision of 1-2 Myr (2σ)⁸, well within the thermal time constant of normal continental crust. In principle, therefore, it should be possible to resolve time differences in mineral growth within a single metamorphic terrain.

Garnet growth in a metasediment from the high-grade part of the metamorphic terrain in Sulitjelma (sample R229; Fig. 1) has recently been studied in detail⁸. This sample records peak metamorphic conditions of $544 \pm 16^\circ\text{C}$ and 7.0 ± 1.0 kbar. Garnet rims and coexisting silicates yield indistinguishable ^{147}Sm - ^{143}Nd and ^{238}U - ^{206}Pb ages of 424.6 ± 1.2 Myr and 423.4 ± 1.7 Myr, respectively. Equally favourable garnet-bearing assemblages exist at lower grades; here we concentrate on a lower-grade rock, located close to the garnet isograd (sample A73; Fig. 1). The sample is a ferromanganiferous metasediment which occurs near the base of the Furulund group, intercalated with strata-bound ore bodies containing Fe, Cu, Zn and S, stratigraphically overlying breccias and pillow lavas of the

Sulitjelma 'ophiolite'. Silicate minerals in this sample include garnet (mole fraction $X_{\text{Fe}} = 0.679$; $X_{\text{Mg}} = 0.055$; $X_{\text{Ca}} = 0.168$), biotite ($\text{Fe}^{2+}/\text{Mg} = 1.17$), chlorite, amphibole, plagioclase (An_{77}), muscovite and quartz. Stilpnomelane and chlorite are also present as retrograde phases. From the method of Powell and Holland^{20,21}, the rim compositions of silicate phases indicate P - T conditions of 6.5 ± 1.0 kbar and $458 \pm 20^\circ\text{C}$, where correlated uncertainties between reactions are minimized for a water activity $a_{\text{H}_2\text{O}} \geq 0.9$. Garnet zoning modelling¹⁰ indicates that there was little change in pressure or temperature during garnet growth. Similar P - T conditions have been derived from other assemblages nearby²², and these are considered to be close to the maximum temperatures experienced at the garnet isograd. At these temperatures significant diffusion of major or trace elements in garnet is unlikely²³.

The precision of any temperature calculated for an exchange equilibrium or discontinuous reaction is no better than $\sim \pm 30^\circ\text{C}$, or a relative error of ± 5 -10%. But the approach of Powell and Holland²⁰ reduces the error because the average temperature is derived from the calculated individual temperatures of a number of independent reactions weighted according to their correlated uncertainty. Temperatures estimated in this way differ by $86 \pm 36^\circ\text{C}$ for the two samples. Iron-magnesium exchange equilibria yield similar temperatures which are independent of $a_{\text{H}_2\text{O}}$, and relatively insensitive to errors in the pressure estimate. For example, application of the garnet-biotite geothermometer of Ferry and Spear²⁴, with recent modifications²⁵⁻²⁷, gives a temperature of $557 \pm 30^\circ\text{C}$ for the high-grade sample R229 and $460 \pm 30^\circ\text{C}$ for the low-grade sample A73 ($\Delta T = 97 \pm 60^\circ\text{C}$). The individual errors are larger than those given above. But the error in the temperature difference between the samples is likely to carry a much smaller uncertainty than $\pm 60^\circ\text{C}$ because of the self-cancellation of errors. In the present case the magnitude of

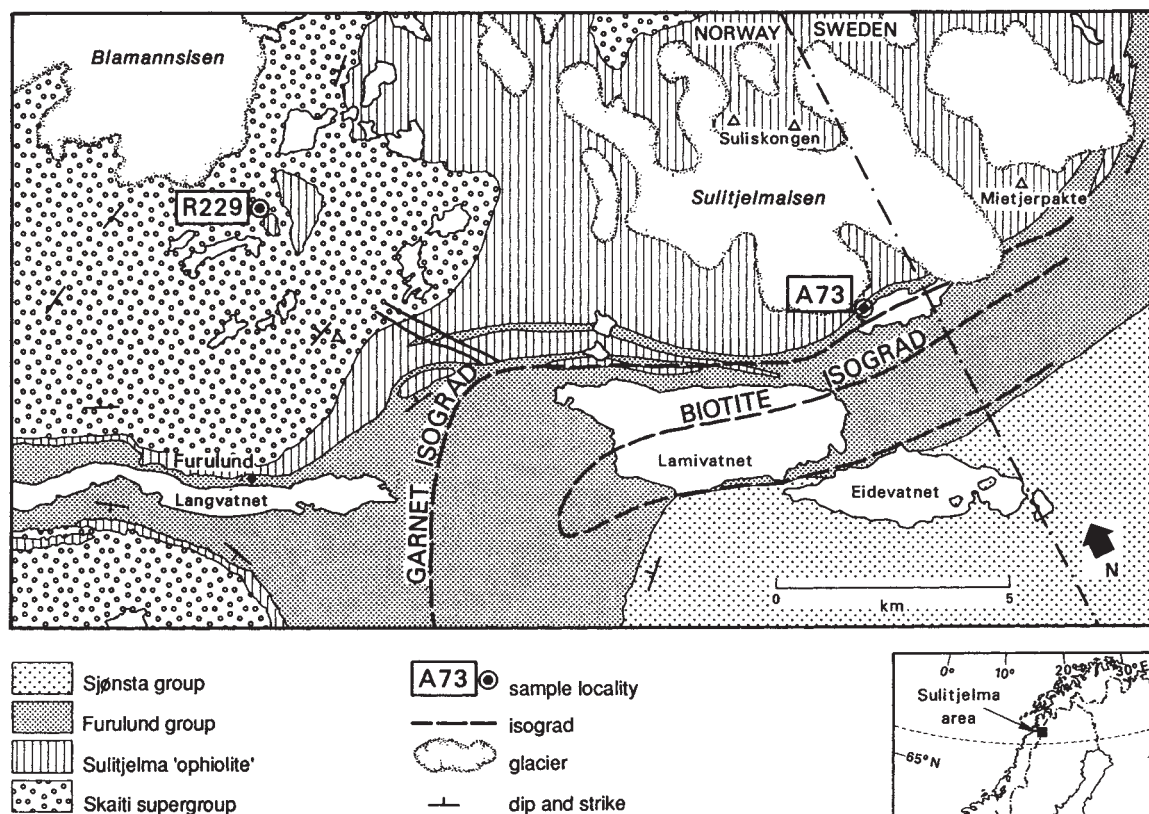


FIG. 1 Geological map of the Sulitjelma area showing sample localities.

TABLE 1 Isotope data for sample A73

Mineral	²⁰⁶ Pb/ ²⁰⁴ Pb	²⁰⁷ Pb/ ²⁰⁴ Pb	²⁰⁸ Pb/ ²⁰⁴ Pb*	⁸⁷ Sr/ ⁸⁶ Sr†	¹⁴³ Nd/ ¹⁴⁴ Nd‡	²³⁸ U/ ²⁰⁴ Pb§		
Garnet	36.568 ± 0.030	16.533 ± 0.016	37.812 ± 0.030	—	0.514805 ± 10	262.94		
Amphibole	18.906 ± 0.014	15.556 ± 0.012	38.422 ± 0.032	0.713486 ± 18	0.512255 ± 6	8.1370		
Biotite	18.359 ± 0.012	15.523 ± 0.012	37.884 ± 0.032	1.017115 ± 92	0.512134 ± 10	0.2693		
Chalcopyrite	18.420 ± 0.014	15.522 ± 0.012	37.839 ± 0.030	0.713508 ± 18	0.512143 ± 6	1.2444		
	⁸⁷ Rb/ ⁸⁶ Sr§	¹⁴⁷ Sm/ ¹⁴⁴ Nd§	[U]	[Pb]	[Rb]	[Sr]	[Sm]	[Nd]
Garnet	—	1.0562	0.2151	0.0649	0.0134	0.2046	2.745	1.572
Amphibole	1.6893	0.1563	1.291	10.13	9.961	16.98	10.23	39.55
Biotite	55.730	0.1161	0.0235	5.488	411.8	21.91	1.912	9.958
Chalcopyrite	1.6931	0.1170	1.066	53.88	1.581	2.689	3.041	15.71

All quoted errors are $2\sigma_m$. Concentrations in p.p.m. by weight. Maximum procedural blanks are 40 pg for Pb and 10 pg for U; 5 pg for Sr and 3 pg for Rb; 5 pg for Nd and 1 pg for Sm.

* Pb isotopic ratios relative to NBS-981 standard, which yields $^{206}\text{Pb}/^{204}\text{Pb} = 16.936 \pm 0.010$; $^{207}\text{Pb}/^{204}\text{Pb} = 15.489 \pm 0.016$; $^{208}\text{Pb}/^{204}\text{Pb} = 36.700 \pm 0.028$ (2σ).

† $^{87}\text{Sr}/^{86}\text{Sr}$ normalized to $^{86}\text{Sr}/^{88}\text{Sr} = 0.1194$; NBS-987 standard yields 0.710231 ± 18 (2σ , external).

‡ $^{143}\text{Nd}/^{144}\text{Nd}$ normalized to $^{146}\text{Sm}/^{144}\text{Nd} = 0.7219$; JN-1 standard yields 0.511129 ± 6 (2σ , external).

§ $^{238}\text{U}/^{204}\text{Pb}$, $^{87}\text{Rb}/^{86}\text{Sr}$ and $^{147}\text{Sm}/^{144}\text{Nd}$ ratios determined to a precision of $\pm 0.3\%$, $\pm 0.25\%$ and $\pm 0.1\%$, respectively.

|| Blank-corrected Pb isotope ratios.

such errors is difficult to assess because of differences in the concentration of non-ideal components in the different assemblages. Nevertheless, the two techniques are in good agreement and indicate that the low-grade sample occurring at the garnet isograd records 'peak' metamorphic temperatures ~ 80 – 90°C lower than in the high-grade sample.

Chemical and isotopic data for garnet, amphibole, biotite and chalcopyrite separated from sample A73 are given in Table 1. Analytical procedures follow those reported previously⁸. Regression of the Sm–Nd data for garnet and coexisting phases yields a best-fit line corresponding to an age of 432.9 ± 1.8 Myr (2σ). These data are shown relative to the best-fit line on a deviation diagram in Fig. 2a. Regression of $^{238}\text{U}/^{204}\text{Pb}$ against $^{206}\text{Pb}/^{204}\text{Pb}$, $^{235}\text{U}/^{204}\text{Pb}$ against $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{204}\text{Pb}$ against $^{206}\text{Pb}/^{204}\text{Pb}$ for the same phases yields concordant ages of 432.1 ± 1.5 Myr, 432.4 ± 6.3 Myr and 433.0 ± 36.0 Myr, respectively (all errors 2σ). The best estimate given by the ^{238}U – ^{206}Pb data (Fig. 2b) is indistinguishable from the ^{147}Sm – ^{143}Nd age obtained for the same minerals.

The P – T – t information obtained here can be combined with that for the higher-grade sample⁸ to provide a precise estimate of the timing of mineral equilibration across this prograde metamorphic terrain. The array of P – T conditions recorded by these two samples indicates a temperature difference of $86 \pm 36^\circ\text{C}$ with little or no change in pressure. The surface P – T array is not, however, directly related to the actual geotherms that existed during metamorphism; it more probably reflects maximum temperatures recorded at different times within a given metamorphic sequence^{1–4}. If regional metamorphism occurs in response to crustal thickening, low-grade rocks are likely to record considerably earlier thermal conditions (cooler geotherms) than will high-grade rocks in the same terrain^{1–4}. In the present case, 'peak' metamorphic conditions at the garnet isograd were attained 8.3 ± 3.0 Myr (Δt , Sm–Nd) or 8.7 ± 3.2 Myr (Δt , U–Pb) before the high-grade sample (see Fig. 3), consistent with increasing burial as the principal cause of metamorphism^{1–4}.

In general, rocks that have experienced crustal thickening

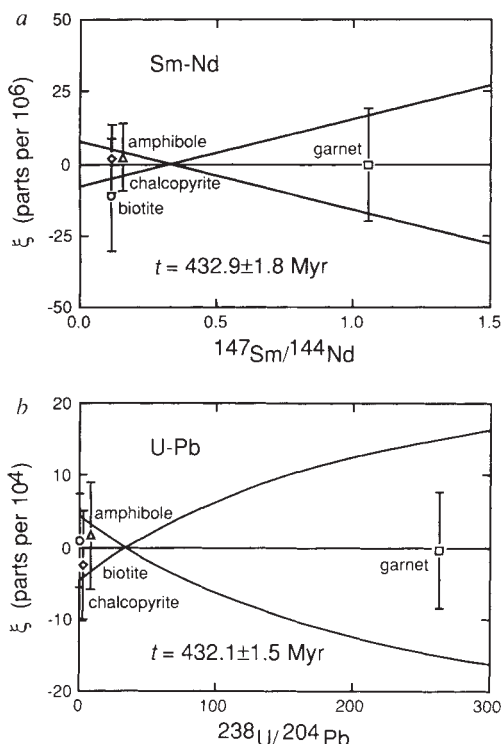


FIG. 2 a, Relative deviation ξ (in parts per 10^6) of measured values from the best-fit isochron, $\xi = [(^{143}\text{Nd}/^{144}\text{Nd})_M / (^{143}\text{Nd}/^{144}\text{Nd})_{\text{BF}} - 1] \times 10^6$ against $(^{147}\text{Sm}/^{144}\text{Nd})_M$ for garnet ($\square$), amphibole (Δ), biotite ($\circ$) and chalcopyrite ($\diamond$) from sample A73, Sulitjelma, Norway. These data define an age of 432.9 ± 1.8 Myr (2σ); mean square weighted deviation is 0.73 and initial $^{143}\text{Nd}/^{144}\text{Nd} = 0.511810 \pm 4$. 2σ error bars are shown for each analysis, and the lines on either side of $\xi = 0$ correspond to 2σ errors derived from the regression analysis. b, Deviation of $^{206}\text{Pb}/^{204}\text{Pb}$ (in parts per 10^4) from the best-fit ^{238}U – ^{206}Pb regression ($\xi = 0$). These data yield an age of 432.1 ± 1.5 Myr (2σ) indistinguishable from the ^{147}Sm – ^{143}Nd age obtained for the same minerals. Mean square weighted deviation is 0.39; initial $^{206}\text{Pb}/^{204}\text{Pb} = 18.339 \pm 0.008$. Note that regression of the data for $^{207}\text{Pb}/^{204}\text{Pb}$ against $^{235}\text{U}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{204}\text{Pb}$ against $^{206}\text{Pb}/^{204}\text{Pb}$ gives concordant ages of 432.4 ± 6.3 Myr and 433.0 ± 36.0 Myr, respectively.

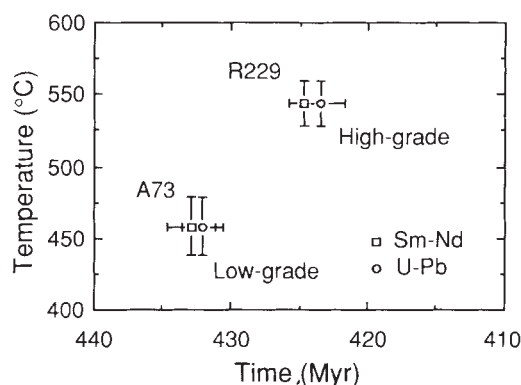


FIG. 3 Temperature-time ($T-t$) diagram showing the conditions and times of mineral equilibration in the Sulitjelma area, northern Norway. Sample A73 occurs close to the garnet isograd (see Fig. 1), and matrix equilibration conditions of $458 \pm 20^\circ\text{C}$ at $6.5 \pm 1.0\text{ kbar}$ are indicated, with little change in P or T during garnet growth. Garnet and coexisting phases yield indistinguishable ^{147}Sm - ^{143}Nd and ^{238}U - ^{206}Pb ages of $432.9 \pm 1.8\text{ Myr}$ and $432.1 \pm 1.5\text{ Myr}$, respectively. Sample R229 is from the high-grade part of this terrain (see Fig. 1), in which final matrix equilibration occurred at $544 \pm 16^\circ\text{C}$ and $7.0 \pm 1.0\text{ kbar}$. Garnet rims and coexisting silicates in this assemblage record indistinguishable ^{147}Sm - ^{143}Nd and ^{238}U - ^{206}Pb ages of $424.6 \pm 1.2\text{ Myr}$ and $423.4 \pm 1.7\text{ Myr}$, respectively⁸. Taken together, these results indicate that mineral equilibration at the garnet isograd occurred $8.3 \pm 3.0\text{ Myr}$ (Δt , Sm-Nd) or $8.7 \pm 3.2\text{ Myr}$ (Δt , U-Pb) before the high-grade sample, consistent with increasing burial as the principal case of metamorphism. Temperatures were calculated using the method of Powell and Holland^{20,21} (all errors 2σ).

spend much of their history within a few degrees of their maximum temperature³. As a consequence although the $P-T$ array presented here does not record any one geotherm that existed during metamorphism, the temperature and time information does allow us to estimate the prograde heating rate. The data indicate average heating rates of $10.4 \pm 12.5^\circ\text{C Myr}^{-1}$ (Δt , Sm-Nd) or $9.9 \pm 5.7^\circ\text{C Myr}^{-1}$ (Δt , U-Pb) over the temperature interval considered here (see Fig. 3). This compares favourably with an independent estimate of the heating rate of $8-9^\circ\text{C Myr}^{-1}$ obtained from differential garnet growth in the high-grade sample. Similar heating rates have been determined for rocks from Vermont ($\sim 5-10^\circ\text{C Myr}^{-1}$)¹⁴ and the eastern Alps ($\sim 3^\circ\text{C Myr}^{-1}$)²⁸.

Our results demonstrate a resolvable time difference in mineral growth across a regional metamorphic sequence. They confirm and quantify the expected behaviour of crustal rocks undergoing thickening where maximum temperatures are attained at different times and different depths in the same metamorphic sequence. □

Received 10 January; accepted 31 March 1992.

- Oxurgh, E. R. & Turcotte, D. L. *Schweiz. Miner. Petrogr. Mitt.* **54**, 641-642 (1974).
- England, P. C. & Richardson, S. W. *J. geol. Soc. Lond.* **134**, 201-213 (1977).
- England, P. C. & Thompson, A. B. *J. Petrol.* **25**, 894-928 (1984).
- Sandiford, M. & Powell, R. *Earth planet. Sci. Lett.* **98**, 154-165 (1990).
- Vogt, T. *Norges. geol. Undersok. Bull.* **121**, 1-560 (1927).
- Mason, R. *Petrology of the Metamorphic Rocks*, 2nd Ed (Unwin Hyman, London, 1990).
- Turner, F. J. *Metamorphic Petrology: Mineralogical, Field and Tectonic Aspects* (McGraw-Hill, New York, 1981).
- Burton, K. W. & O'Nions, R. K. *Earth planet. Sci. Lett.* **107**, 649-671 (1991).
- Essene, E. *Evolution of Metamorphic Belts, Geol. Soc. Spec. Publ.* **43**, 1-44 (1989).
- Spear, F. S. & Selverstone, J. *Contrib. Mineral. Petrol.* **83**, 348-357 (1983).
- Cohen, A. S., O'Nions, R. K., Siegenthaler, R. & Griffin, W. I. *Contrib. Mineral. Petrol.* **98**, 303-311 (1988).
- Jagoutz, E. *Geochim. cosmochim. Acta* **52**, 1285-1293 (1988).
- Mezger, K., Hanson, G. N. & Bohlen, S. R. *Contrib. Mineral. Petrol.* **101**, 136-148 (1989).
- Christensen, J. N., Rosenfeld, J. L. & DePaolo, D. J. *Science* **244**, 1465-1469 (1989).
- Vance, D. & O'Nions, R. K. *Earth planet. Sci. Lett.* **97**, 227-240 (1990).
- Harrison, W. J. & Wood, B. J. *Contrib. Mineral. Petrol.* **72**, 145-155 (1980).
- Coghlan, R. A. N. thesis, Brown Univ., Rhode Island (1990).
- Cygan, R. T. & Lasaga, A. C. *Am. J. Sci.* **285**, 328-350 (1985).
- Elphick, S. C., Ganguly, J. & Loomis, T. P. *Contrib. Mineral. Petrol.* **90**, 36-44 (1985).
- Powell, R. & Holland, T. J. B. *J. met. Geol.* **6**, 173-204 (1988).
- Holland, T. J. B. & Powell, R. J. *met. Geol.* **8**, 89-124 (1990).
- Boyle, A. P. & Westhead, R. K. *J. met. Geol.* (in the press).
- Chakraborty, S. & Ganguly, J. *Adv. phys. Geochem.* **8**, 120-175 (1991).

- Ferry, J. M. & Spear, F. S. *Contrib. Mineral. Petrol.* **66**, 113-117 (1978).
- Ganguly, J. & Saxena, S. K. *Am. Min.* **69**, 88-97 (1985).
- Ganguly, J. & Saxena, S. K. *Am. Min.* **70**, 1320 (1986).
- Hackler, R. T. & Wood, B. J. *Am. Min.* **74**, 994-999 (1990).
- Christensen, J. N., Rosenfeld, J. L., DePaolo, D. J. & Selverstone, J. *EOS* **72**, 530 (1991).

ACKNOWLEDGEMENTS. We thank R. Powell, J. Selverstone and J. Ganguly for reviews. J. Ganguly is also thanked for a recent formulation of the garnet-biotite geothermometer. This research was supported by the NERC and the Royal Society.

Female swallow preference for symmetrical male sexual ornaments

Anders Pape Møller*

Department of Zoology, Uppsala University, Box 561, S-751 22 Uppsala, Sweden

MANY secondary sexual characters are supposed to have evolved as a response to female choice of the most extravagantly ornamented males¹, a hypothesis supported by studies demonstrating female preferences for the most ornamented males²⁻⁵. Comparative studies of elaborate feather ornaments in birds have shown that (1) ornaments have larger degrees of fluctuating asymmetry⁶ (small, random deviations from bilateral symmetry caused by an inability of individuals to cope with environmental and genetic stress during development of a character⁷) than other morphological traits, and (2) the degree of fluctuating asymmetry is often negatively related to the size of the ornament⁶. The negative relationship between ornament asymmetry and size suggests that ornament size reliably reflects male quality because the largest secondary sex traits demonstrate the least degree of fluctuating asymmetry. I manipulated tail length and tail asymmetry independently in male swallows (*Hirundo rustica*) to determine whether ornament size or asymmetry were used as cues in mate choice. Male swallows with elongated, symmetric tails mated earlier, and enjoyed larger annual reproductive success than did males with shortened tails and increased asymmetry. Females therefore prefer large as well as symmetric ornaments, which suggests that females in their mate choice use ornament asymmetry and size as reliable indicators of male quality.

I simultaneously and independently manipulated ornament size and asymmetry in swallows, which are small (about 20 g), insectivorous monogamous passerines that feed on the wing and usually nest in small colonies. Sexual size dimorphism is slight for most external body traits, with the exception of the outermost long tail feathers (tail length), which in my study population are on average 20% longer in males than in females³. Fluctuating asymmetry in male tail length is on average 2.10 mm (s.e.m., 0.18; range, 0-49; $N = 517$), which is much larger than in other morphological characters⁸. There is a strong negative relationship between ornament asymmetry and ornament size in male swallows⁸. Naturally and experimentally long-tailed males are more likely to attract a mate than are short-tailed males, and, among those that do attract a mate, achieve one more quickly than do short-tailed males^{3,8}.

I captured unmated swallows in mist nets at Kraghede ($57^\circ 12' \text{N}$, $10^\circ 00' \text{E}$), Denmark, during the period 10 May to 15 June 1991 shortly after their arrival and provided them with an individual combination of colour rings and a numbered aluminium ring. The swallows breed on farms either solitarily or in colonies of up to 50 pairs. A number of standard morphological measurements were taken, including the length of the

* Present address: The Galton Laboratory, Department of Genetics and Biometry, University College of London, 4 Stephenson Way, London NW1 2HE, UK.

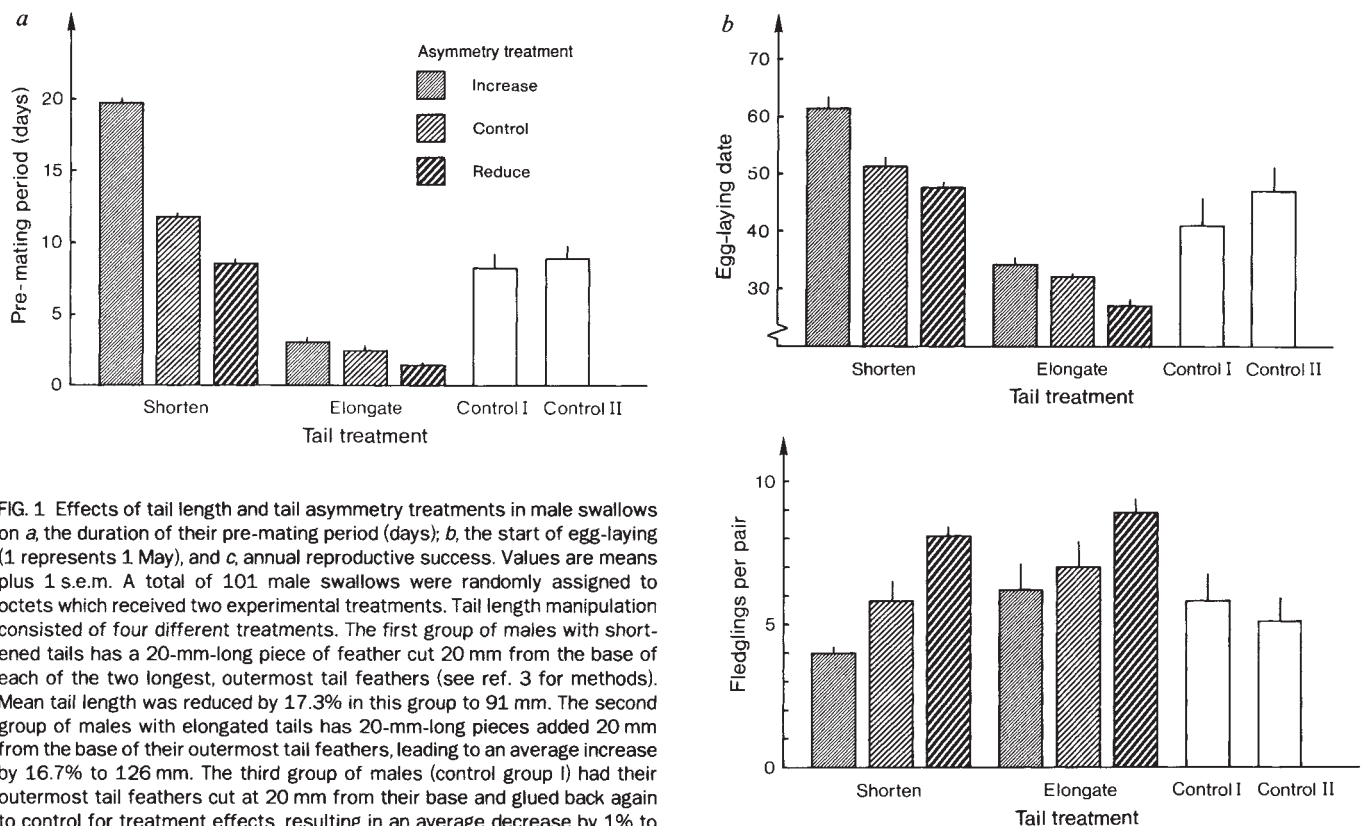


FIG. 1 Effects of tail length and tail asymmetry treatments in male swallows on *a*, the duration of their pre-mating period (days); *b*, the start of egg-laying (1 represents 1 May), and *c*, annual reproductive success. Values are means plus 1 s.e.m. A total of 101 male swallows were randomly assigned to octets which received two experimental treatments. Tail length manipulation consisted of four different treatments. The first group of males with shortened tails has a 20-mm-long piece of feather cut 20 mm from the base of each of the two longest, outermost tail feathers (see ref. 3 for methods). Mean tail length was reduced by 17.3% in this group to 91 mm. The second group of males with elongated tails has 20-mm-long pieces added 20 mm from the base of their outermost tail feathers, leading to an average increase by 16.7% to 126 mm. The third group of males (control group I) had their outermost tail feathers cut at 20 mm from their base and glued back again to control for treatment effects, resulting in an average decrease by 1% to 105 mm. The fourth group (control group II) was captured, but their tail feathers were not manipulated, and had an average length of 106 mm. The second treatment consisted of manipulating the degree of asymmetry of the two outermost tail feathers by either (1) increasing asymmetry by 20 mm (to an average of 23 mm) without changing the mean length of the two outermost tail feathers following the first treatment; (2) decreasing asymmetry to 0 mm; or (3) keeping males as controls without changing the original level of asymmetry, on average 3 mm. The two experimental treatments of the tails of male swallows were effective as evidenced from the highly significant differences in tail length in relation to tail length treatment

(one-way ANOVA: $F = 35.15$, d.f. = 7, 88, $P < 0.0001$), although there was no difference before treatment ($F = 0.73$, d.f. = 7, 88, NS). Similarly, the asymmetry treatment was effective as the groups of males receiving the increased asymmetry treatment really were more asymmetric in their tail ornaments than the controls, which were more asymmetric in their tails than males in the group receiving the decreased asymmetry treatment (one-way ANOVA: $F = 98.89$, d.f. = 7, 88, $P < 0.0001$), although there was no difference before treatment ($F = 0.42$, d.f. = 7, 88, NS). Sample sizes were 12, 10, 11, 12, 12, 12, 14 and 13, respectively, for the eight groups.

two outermost elongated tail feathers⁹. I tested experimentally whether length and fluctuating asymmetry of male tails independently affected mate choice. The first treatment consisted of elongating, shortening or control treating the length of the two outermost tail feathers. The second simultaneous treatment consisted of manipulation of the degree of fluctuating asymmetry of the two outermost tail feathers of males having their tail feathers elongated or shortened. Male swallows were thus randomly assigned to one of eight different groups (shortened tails with increased, control or decreased asymmetry, elongated tails with increased, control or decreased asymmetry, tails just cut and glued back again, or tails unmanipulated) with the restriction that males were assigned to all eight groups before assignment of males to a new octet (Fig. 1).

The effects of experimental treatments were assessed at different stages of the reproductive cycle. Five male swallows disappeared after being released, whereas all other males subsequently acquired a mate. The duration of the pre-mating period of male swallows was significantly affected by both experimental treatments (Fig. 1). There was no marked effect of the tail cutting and glueing procedure *per se*, because males with cut and glued tails took as long a time to acquire a mate as did males without tail treatment (Fig. 1). I therefore analysed the effects of the tail length treatment and the tail asymmetry treatment on reproductive parameters in a two-way analysis of variance by using the six groups of males receiving both treatments. The duration of the pre-laying period was affected by tail treatment, tail asymmetry treatment and their interaction

(Table 1). Long-tailed male swallows took longer to acquire a mate if asymmetric rather than symmetric, but the difference between asymmetric and symmetric short-tailed males was much larger (Fig. 1). The effects of treatments on the duration of mate acquisition subsequently affected the time of egg laying, which was also affected by both treatments and their interaction (Table 1). Clutch size was larger among females mated to males with elongated tails compared with males with shortened tail feathers (Table 1). But brood size at fledging was not statistically significantly affected by experimental treatments (Table 1). Total annual reproductive success of males differed significantly in relation to the manipulations of their outermost tail feathers (Fig. 1 and Table 1). Male swallows with elongated tail feathers fledged more young per season than did males with shortened tail feathers, but there was also an effect of the asymmetry treatment, because males with symmetric tails produced more offspring than controls that produced more offspring than males with asymmetric tails (Fig. 1). There was a significant tail length by asymmetry interaction; the difference in annual reproductive success between long-tailed and short-tailed males was larger if they had asymmetric rather than symmetric tails, with individuals receiving the control asymmetry treatment in between (Fig. 1). In conclusion, both tail length and tail asymmetry influenced sexual selection and subsequent annual reproductive success in male swallows.

The results of the experimental treatment could be due to male-male competition, female choice, or both¹. I tested whether treatments affected the intensity of male-male competition by

TABLE 1 The effects of tail length treatment and tail asymmetry treatment of male swallows on their timing of mate acquisition and reproductive parameters

Reproductive parameter	Tail-length treatment	Asymmetry treatment	Interaction
Pre-mating period	278.32***	207.62***	241.67***
Date of egg laying of first clutch	331.55***	25.47***	3.86*
First clutch size	8.03**	0.13	0.34
First brood size at fledging	0.15	0.38	0.11
Second clutch size	2.78	0.10	5.23*
Second brood size at fledging	3.21	0.65	5.66*
Total number of fledglings	5.67*	11.30***	5.76**

F-values are from two-way ANOVAs. The degrees of freedom for the pre-mating period, date of laying of first clutch, first clutch size, first brood size at fledging and total number of fledglings are 1, 55 for tail length treatment and 2, 55 for asymmetry treatment and the interaction. The degrees of freedom for second clutch size and second brood size at fledging are 1, 33 for tail length treatment and 2, 33 for asymmetry treatment and the interaction. The duration of the pre-mating period and the time of start of egg laying were log-transformed before calculations in order to obtain roughly normal distributions. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

recording the rate of aggressive male-male interactions by focal male swallows during one-hour morning (05:00–12:00) observation periods on the first three days following treatment. There was no effect of tail length manipulation, asymmetry manipulation, or their interaction on the rate of male-male aggressive encounters (two-way ANOVA: tail length treatment, $F = 0.03$, d.f. = 1, 90, NS; asymmetry treatment, $F = 2.50$, d.f. = 2, 90, NS; interaction, $F = 1.48$, d.f. = 2, 90, NS). Males almost always won aggressive encounters on their own territories.

Female swallows visit up to several different males before making their mate choice^{3,9}. I tested whether treatments affected mate choice by recording the number of female visits at the territories of focal male swallows during three one-hour morning observation periods as already described. Males that experienced difficulties in acquiring a mate were predicted to be visited and rejected by more females than preferred male swallows. There was an effect of tail length treatment and asymmetry treatment on whether male swallows were visited by more than one female, but their interaction was not significant (log-likelihood ratio test: tail length treatment, $G^2 = 7.09$, d.f. = 1, $P < 0.01$; asymmetry treatment, $G^2 = 22.60$, d.f. = 2, $P < 0.01$; interaction, $G^2 = 3.25$, d.f. = 2, NS). This suggests that the experiment affected male reproductive success as determined by female choice rather than by male-male competition for access to females.

Fluctuating asymmetry in the outermost tail feathers of males could directly affect manoeuvrability and thus ability to capture prey¹⁰. I determined whether this was the case by measuring offspring quality in relation to experimental treatment. Offspring quality measured in terms of tarsus length (a skeletal character kept throughout life) and body mass on day 15 of the nestling period was unrelated to experimental treatments (two-way ANOVA: tarsus length, tail treatment, $F = 1.18$, d.f. = 1, 90, NS; asymmetry treatment, $F = 0.98$, d.f. = 2, 90, NS; interaction, $F = 0.77$, d.f. = 2, 90, NS; body mass, tail treatment, $F = 1.12$, d.f. = 1, 90, NS; asymmetry treatment, $F = 0.74$, d.f. = 2, 90, NS; interaction, $F = 0.14$, d.f. = 2, 90, NS). There was therefore no direct effect of experimental tail length or tail asymmetry of male swallows on the quality of their offspring.

Male swallows could develop one long and one short outermost tail feather during moult, rather than two relatively short, symmetric feathers, if only ornament size, and not symmetry of ornaments, had been important in sexual selection. But the

results of my experiment demonstrate that males simultaneously have to optimize both ornament size and ornament asymmetry because both directly affects sexual selection. Fluctuating asymmetry is an epigenetic measure of the ability of individuals to cope with environmental and genetic stress^{1,11,12}, and the negative relation between the degree of asymmetry and ornamentation in male swallows in particular and in many birds in general therefore suggests that the expression of ornaments reliably reflects male quality in terms of the ability of males to cope with their environment (ref. 6 and A.P.M. and A. Pomiankowski, manuscript in preparation). This is consistent with 'good genes' models of sexual selection^{13,14}, which predict that females choose males with the most extravagant ornaments because they reliably reflect the genetic quality of the male. A long-lasting debate between proponents of different groups of models of mate choice has generated much controversy¹⁵. The general patterns of fluctuating asymmetry in sex traits in birds in relation to the size of the ornaments and the results of the present experiment support the 'good genes' models of mate choice. But this does not exclude the possibility that other models of mate choice, which predict different patterns of fluctuating asymmetry (A.P.M. and A. Pomiankowski, manuscript in preparation), can work under other conditions. □

Received 27 January; accepted 17 March 1992.

1. Darwin, C. *The Descent of Man and Selection in Relation to Sex* (Murray, London, 1871).
2. Andersson, M. *Nature* **299**, 818–820 (1982).
3. Møller, A. P. *Nature* **332**, 640–642 (1988).
4. Smith, H. G. & Montgomerie, R. D. *Behav. Ecol. Sociobiol.* **28**, 195–201 (1990).
5. Petrie, M., Halliday, T. R. & Sanders, C. *Anim. Behav.* **41**, 323–331 (1990).
6. Møller, A. P. & Höglund, J. *Proc. R. Soc. B* **245**, 1–5 (1991).
7. Ludwig, W. *Das Rechts-Links Problem im Tierreich und beim Menschen* (Springer, Berlin, 1932).
8. Møller, A. P. *Anim. Behav.* **40**, 1185–1187 (1990).
9. Møller, A. P. *Anim. Behav.* **39**, 458–465 (1990).
10. Møller, A. P. *Proc. R. Soc. B* **243**, 59–62 (1991).
11. Palmer, A. R. & Strobeck, C. A. *Rev. Ecol. Syst.* **17**, 391–421 (1986).
12. Parson, P. A. *Biol. Rev.* **65**, 131–145 (1990).
13. Trivers, R. L. in *Sexual Selection and the Descent of Man, 1871–1971* (ed. Campbell, B.) 136–179 (Aldine, Chicago, 1972).
14. Zahavi, A. *J. theor. Biol.* **53**, 205–214 (1975).
15. Kirkpatrick, M. & Ryan, M. J. *Nature* **350**, 33–38 (1990).

ACKNOWLEDGEMENTS. I thank A. Pomiankowski and S. Ulfstrand for comments on the manuscript. The research was supported by a grant from the Swedish NSRC and the British NERC.

Long-term potentiation is associated with increases in quantal content and quantal amplitude

Dimitri M. Kullmann & Roger A. Nicoll*

Departments of Pharmacology and Physiology, School of Medicine, University of California, San Francisco, San Francisco, California 94143–0450, USA

LONG-TERM potentiation (LTP) of synaptic transmission in CA1 neurons of the hippocampus, elicited by the conjunction of presynaptic firing and postsynaptic depolarization, is an important model of plasticity, which may underlie memory storage^{1–3}. Although induction of LTP takes place in the postsynaptic cell^{4–7}, it is not clear whether it is expressed through an enhancement of transmitter release^{8–12} or through an increased postsynaptic response to the same amount of transmitter^{13–16}. Analysis of the trial-to-trial amplitude fluctuations of synaptic signals, that is quantal analysis, gives an important insight into the probabilistic mechanisms of transmission, although attempts to apply it to the mode of expression of LTP have so far yielded inconsistent results^{9–12,15}, at least in part because they have relied on models

* To whom correspondence should be addressed.

of transmitter release that have not been confirmed experimentally¹⁷⁻¹⁹. Here we report clear evidence for quantal fluctuation in a subset of cells. Induction of LTP in these cells causes abrupt increases in either quantal content or quantal amplitude, or both. This shows that two different mechanisms can underlie the maintenance of LTP.

For quantal analysis to be applied to synaptic transmission, the different release sites must give rise to very similarly sized postsynaptic signals so that the evoked signal fluctuates among discrete levels separated by a constant increment. This is difficult to reconcile with the very broad and skewed amplitude distribution of spontaneous excitatory postsynaptic currents (e.p.s.cs), recorded in CA1 neurons using whole-cell patch-clamping in the presence of tetrodotoxin^{16,20}. The possibility, however, remains that this variability is due principally to nonuniformity across synapses, rather than nonstationarity in time at each synapse. Among a large sample of evoked e.p.s.cs studied in different experiments there would then be a subset where the release events were roughly of the same amplitude, and where the quantal model therefore held²¹. Rather than assume that every synaptic input we studied conformed to the quantal model, we instead examined each case for quantal fluctuation, and only if this was observed did we look at the effect of LTP induction on the quantal parameters. The mean quantal content, m , and the mean quantal amplitude, Q , were estimated not by constraining the fluctuation pattern to agree with an *a priori* probabilistic model of transmitter release such as a binomial distribution,

but instead from the peaks, if any, in the e.p.s.c. amplitude distribution.

We used maximum entropy noise deconvolution to separate the underlying statistical features of the amplitude fluctuation from the effects of random sampling and background noise. Conventional deconvolution methods rely on optimizing the accuracy of fit or likelihood function defined by two distributions: the observed synaptic signal amplitude distribution, and the reconvolution of the noise distribution with the underlying fluctuation pattern which is being sought²². Because the amplitude distribution is finitely sampled, this approach is biased towards finding a number of discrete amplitude levels, some of which reflect sampling artefacts rather than true features of the data. The deconvolution method thus tends to overfit the data, and many alternative solutions can be proposed which give almost as good fits. We attempted to correct this bias by maximizing the entropy, or smoothness, of the optimization solution, at the same time as maximizing the likelihood function describing the accuracy of fit to the observed data. The balance between entropy and likelihood was chosen to find a unique solution which is the most featureless distribution still compatible with the data (the maximum entropy noise deconvolution, MEND, solution; Fig. 1 legend).

This approach gives a conservative interpretation of the data, because it does not rely on the constraints imposed by the quantal assumption, nor does it maximize the accuracy of fit beyond what should be expected. This has been confirmed with

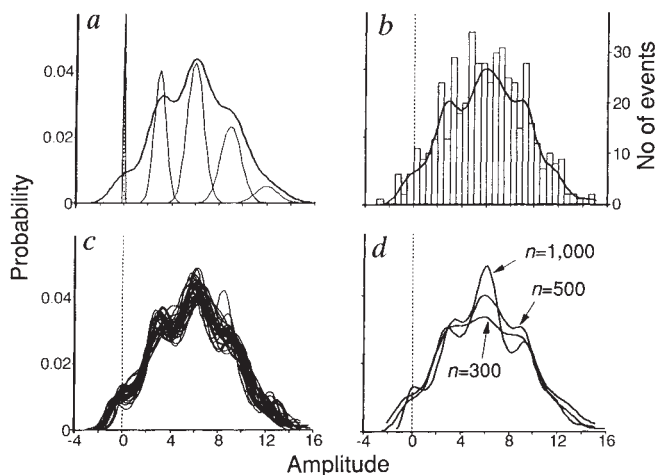


FIG. 1 Examples of Monte Carlo simulations of the MEND approach. A quantal model was simulated, with four release sites with probability = 0.5, quantal amplitude = 3, and quantal coefficient of variation = 1.7% (that is, the standard deviation of the component at 3 was 0.5). Gaussian noise with standard deviation of 1 was added. *a*, The underlying components are shown, together with their sum after convolution with the noise function. *b*, 500 points were randomly sampled from this distribution. The MEND solution drawn as a continuous line through the resulting histogram shows peaks at about 3, 6 and 9. *c*, Results for 29 different random samples of the same size drawn from the same population density function. Although in many cases the peaks at 0, 9 and 12 were not resolved, in no case was an additional spurious peak identified. *d*, Comparison of results obtained for random samples of different sizes. As the number of data points is reduced, the peaks gradually disappear. Conversely, the resolution of the peaks is improved as the data sample is increased. In neither case are spurious peaks seen.

METHODS. The data (experimental or simulated e.p.s.c. amplitudes) are contained in a frequency histogram $\mathbf{a}$ ($a_i, i = 1, \dots, k, \dots, n$; that is, a_i is the number of trials giving rise to an amplitude of $v_i \pm \delta/2$, where δ is the binning interval, and k indicates the bin at 0 pA). The noise distribution $\mathbf{g}$ is defined over the same interval (for the purpose of the simulations it was assumed to be a gaussian distribution with standard deviation of 1, but in the analysis of experimental data, it was obtained by sampling the background noise). The required solution is a probability density function $\mathbf{f}$

($f_j, j = 1, \dots, n$), also defined over the same range as $\mathbf{a}$ and $\mathbf{g}$. The goal is to find $\mathbf{f}$ such that the convolution of $\mathbf{f}$ and $\mathbf{g}$ gives the maximum likelihood fit to $\mathbf{a}$, while simultaneously maximizing the entropy of $\mathbf{f}$. The likelihood function describing this fit is given in its familiar logarithmic form by $L = \sum_i a_i \log \sum_j f_j g_{i-j+k}$. This can be maximized with respect to $\mathbf{f}$ by applying the expectation-maximization algorithm^{24,25}: starting with an arbitrary distribution, $\mathbf{f}$ is iteratively optimized. At each cycle, each element f_j is updated to f'_j using the recursion:

$$f'_j = \frac{1}{\sum_i a_i} \sum_i \frac{f_j g_{i-j+k}}{\sum_j f_j g_{i-j+k}}$$

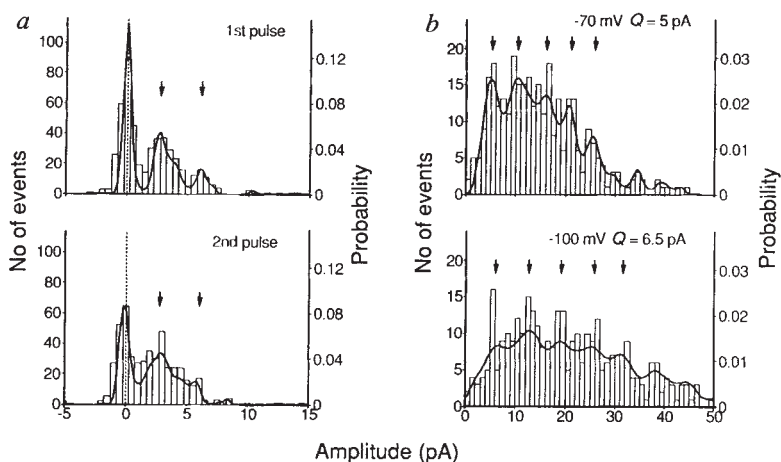
Maximizing the entropy of $\mathbf{f}$ is trivial: because $\mathbf{f}$ is a probability distribution which must sum to 1, this is simply done by setting $f_j = 1/n$ for $j = 1, \dots, n$. By analogy with the method proposed by Gull and Daniell²⁶ to deal with similar problems in astronomy, likelihood and entropy are simultaneously maximized by applying Lagrange's method of unknown multipliers. The recursion is modified to:

$$f'_j = \left(\frac{1}{\sum_i a_i} \sum_i \frac{f_j g_{i-j+k}}{\sum_j f_j g_{i-j+k}} \right)^\lambda \left(\frac{1}{n} \right)^{1-\lambda}$$

where λ can take a value between 0 (which gives the maximum entropy solution) and 1 (which gives the maximum likelihood solution). Starting with a flat distribution and an intermediate value of λ , say 0.5, the recursion is applied until it passes a convergence threshold; the solution $\mathbf{f}$ is normalized to sum to 1, reconvolved with the noise distribution $\mathbf{g}$, and compared to the e.p.s.c. amplitude histogram $\mathbf{a}$ with a goodness-of-fit test. If this fails at a predetermined level of significance, λ is increased, and the recursion is applied again until the goodness-of-fit criterion is satisfied. (The algorithm is described for the case where $\mathbf{a}$, $\mathbf{g}$ and $\mathbf{f}$ are discrete functions defined over the same space, so the χ^2 test can be used to assess the goodness-of-fit, but it can equally be applied to unbinned data, in which case the Kolmogorov-Smirnov test performs the same function. All the results presented were verified by using both methods. The illustrated histograms were rebinned with a binwidth at least twice as large to help comparison with the MEND solutions.) The precise trade-off between likelihood and entropy is determined by the level of significance at which the iterations are allowed to stop. Following ref. 26, we aimed for a χ^2 value equal to the number of degrees of freedom, or equivalently a probability of 0.5 that a better fit could have been observed. An even more stringent test to identify peaks in the observed data would be to aim for a probability of 0.05, which would maximize entropy further at the expense of likelihood and would place the solution just at the borderline for rejection as being too smooth to account for the observed data at the 5% significance level. We have applied both criteria to the results described here, and found very little difference in outcome.

FIG. 2 Effects of varying quantal content and quantal amplitude on histograms and MEND solutions. *a*, Paired-pulse facilitation yields an increase in quantal content with no change in quantal amplitude levels. Amplitude histograms and MEND solutions are shown for (top) first and (bottom) second e.p.s.c.s recorded from a CA1 cell with an interpulse interval of 50 ms. Arrows in this and subsequent figures indicate prominent peaks in the MEND solution. Only the peaks near the origin are indicated because small inequalities between the events making up the e.p.s.c. are more likely to cause deviation from the strictly quantal model as the number of release events increases. (Small inflexions on the peaks are ignored for the same reason.) *b*, Quantal amplitude changes with holding potential. Amplitude histograms and MEND solutions obtained in the same cell at (top) -70 mV and (bottom) -100 mV. (Noise standard deviation (n.s.d.) was 1.8 pA at -70 mV and 2.0 pA at -100 mV.)

METHODS. Experiments were carried out on $450\text{-}\mu\text{m}$ thick transverse hippocampal slices obtained from 4–5-week-old guinea pigs, using standard techniques^{16,27}. The superfusing medium, unless otherwise stated, contained (in mM) NaCl, 119; KCl, 2.5; MgCl_2 , 1.3; CaCl_2 , 2.5; NaHCO_3 , 26.2; NaH_2PO_4 , 1; glucose, 11; picrotoxin, 0.1; and was bubbled with 95% O_2 /5% CO_2 (pH 7.4). CA3 was cut away and recordings were made from CA1 cells at 23 or 30°C using the blind whole-cell patch-clamp technique^{28,29}. The pipette solution contained (in mM): Cs-glucuronate, 122.5; CsCl, 17.5; HEPES, 10; EGTA, 0.2; NaCl, 8; Mg-ATP, 2 (or 4); GTP, 0.3 (pH 7.2, osmolality 290–295 mOsm). With whole-cell recording LTP typically could not be elicited after 20–30 min of recording. In a number of cells the nystatin perforated patch technique³⁰ was used so that more trials could be obtained during the control period. Stimuli were delivered through fine bipolar stainless steel electrodes positioned in stratum radiatum close to stratum pyramidale, but as far as possible from the recording site in an attempt to limit the number of fibres carrying the incoming volley. The stimulus strength was graded to obtain small response with frequent apparent failures of transmission. In cases with clear failures of transmission it was verified that small changes in stimulus strength had no effect on the failure rate, but it could not be held with certainty that single-fibre e.p.s.c.s were recorded. In most cases paired pulses were delivered with an interpulse interval of 50 ms, and a trial repetition rate of between 1 and 0.25 Hz. This was kept high to collect as many data points as possible, but varied between slices in response to the variable tendency for the e.p.s.c. to run down. Results were accepted if the response remained stable during a baseline period longer than 10 min. In some experiments, single pulses were delivered through a second electrode acting as a control pathway. This was switched off during the depolarization



used to elicit LTP to check that the potentiation was restricted to the test pathway. Cells were clamped at -80 or -90 mV using an Axopatch 1D amplifier, and the clamp current was filtered at 2 kHz, digitized at 2 or 3.33 kHz, and stored on an IBM AT compatible computer equipped with a Labmaster A/D board. The data records consisted of the evoked synaptic currents, together with a baseline noise period and the current transient in response to a 1 mV step command, which was used to monitor the series resistance. Results were discarded if the series resistance changed by more than 15%, or if there was not an immediate increase of more than 30% in the mean e.p.s.c. amplitude after the pairing procedure. LTP ranged from 35 to 430%, and the analysis was restricted to results obtained in the first 30 min after induction. To measure the e.p.s.c. amplitude, two periods were chosen, corresponding to the peak of the average e.p.s.c. time course and a baseline immediately before the e.p.s.c., and the difference between these windows was computed for every trial. The noise amplitude was similarly measured by repeating the measurement during part of the data records when the e.p.s.c.s were not evoked, and this was repeated several times for each trial to obtain a large sample. The e.p.s.c. amplitudes were binned to give a frequency histogram *a*, defined in the legend to Fig. 1. The background noise distribution, *g*, was obtained using the same binning interval. Because the noise was sampled many more times than the e.p.s.c., *g* could be known with almost arbitrary accuracy and was fitted by a continuous unimodal function²⁴.

extensive Monte Carlo simulations (Fig. 1): when the underlying quantal amplitude levels were widely spaced in relation to the noise, then the correct amplitudes were resolved with no additional components. By contrast, when the underlying amplitude levels were very close together, or merged into a continuous distribution, or if quantal variability was simulated, then the solution was smooth, and in contrast to results obtained by maximizing likelihood alone, there were no spurious inflexions which might suggest an incorrect quantal amplitude estimate (Fig. 1c). Moreover, if for a given underlying model the sample size was gradually reduced, the MEND solution became progressively flatter, until all inflexions were lost, and again no spurious peaks were seen (Fig. 1d). In an experimental situation, therefore, if the solution has clear peaks occurring at equal increments, with the first peak at 0, then this is strong evidence that the quantal model is valid. This was indeed the case in 27 of 98 experiments where e.p.s.c.s elicited by minimal stimulation of stratum radiatum were recorded with low-noise patch-clamp techniques. The remainder of the results fell into two categories: either the solution was smooth, with no clear inflexions, even though inspection of the raw amplitude histogram sometimes showed what appeared to be quantal peaks; or there was a large number of peaks that were not equally spaced. We interpret the first results as arising from cases where the quanta were too small or too variable to be resolved, and the second as cases where the quantal model was not satisfied because of nonuniformity between the active synapses.

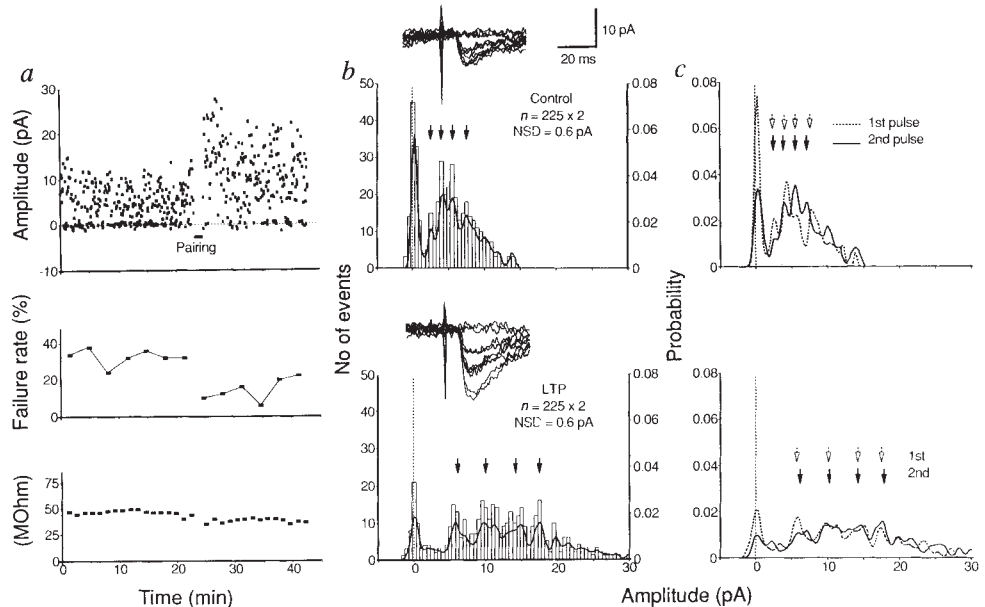
We used paired pulse facilitation to verify experimentally that

the quantal peaks were correctly determined. By analogy with other synapses²³, this should be mediated presynaptically by an increase in quantal content with no change in quantal amplitude. We compared MEND solutions for the first and second e.p.s.c.s recorded with paired pulses separated by 50 ms. The underlying fluctuation patterns were similar for both pulses, with peaks occurring in the same positions, except that for the second pulse there were fewer trials giving rise to failures and lower amplitude peaks, and more trials resulting in larger amplitude peaks (Fig. 2a). This indicates that the estimates of the quantal amplitude peaks were consistent within an experiment, and further implies that a change in quantal content was correctly identified. Paired pulses were used in 11 of the experiments on LTP, and in each of these the amplitude peaks showed good agreement between the first and second pulses.

We next verified that a change in quantal amplitude could be detected, and would not be misinterpreted as a change in quantal content. This was achieved by recording e.p.s.c.s at two different holding potentials: -70 and -100 mV ($n=5$). At the more negative potential the greater driving force should result in a greater separation between the quantal amplitude levels. This was indeed the case: in Fig. 2b, taken from a representative cell, the average separation between the first four peaks was about 5 pA at -70 mV, and 6.5 pA at -100 mV, agreeing well with the difference expected from a reversal potential around 0 mV.

What effect does LTP have on the failure rate and the quantal peaks? Figure 3a, top, shows a graph in which the amplitude of each e.p.s.c. is plotted over time. The response clearly

FIG. 3 LTP obtained in a CA1 cell recorded with the nystatin perforated-patch technique. *a* (top), The e.p.s.c. amplitude for the first of the two pulses before and after LTP induced by pairing stimulation with postsynaptic depolarization. *a* (middle), The failure rate, estimated as the percentage of trials resulting in a response < 2 pA in amplitude, is plotted for successive averages of 50 sweeps. An increase in quantal content is indicated by the drop in failure rate. (A similar change was observed for the second pulse.) *a* (bottom), Access resistance monitored throughout the experiment. *b*, Amplitude histograms and MEND solutions obtained by summing together the data from the first and second pulses for the last 225 sweeps before (top), and the first 225 sweeps after LTP induction (bottom). Sample consecutive records for the first pulse are shown. The first four prominent peaks in the MEND solutions, not counting the peak at 0 pA, are indicated by arrows. The increased separation after LTP indicates an increase in quantal amplitude. The reduction in the height of the peak at 0 pA confirms that the quantal content has also increased. *c*, MEND solutions for the e.p.s.c.s elicited by the first and second pulses calculated separately for the same control (top) and LTP (bottom) periods as in *b*.



fluctuates between failures of transmission and responses of variable size. LTP was induced by depolarizing the cell to 0 mV for 2 min while continuing stimulation at the same intensity and rate (total of 60 impulses). Immediately after the pairing there was a decrease in the incidence of failures (Fig. 3*a*, middle). This is also clearly shown in the histograms obtained before and during LTP (Fig. 3*b*) indicating that the quantal content of the e.p.s.c. increased after LTP. There was also an increase in quantal amplitude, shown by the increased separation between the prominent peaks in the amplitude histograms obtained by pooling together the e.p.s.c.s elicited by the first and second pulses. The MEND solutions, calculated from unbinned data, confirmed that the peaks arose from clustering in the underlying fluctuation pattern which could not be accounted for by the background noise. The standard deviation of the noise, moreover, did not change appreciably between the two periods. As a further test that the underlying fluctuation patterns were correctly identified, the MEND solutions were calculated for the first and second pulses separately (Fig. 3*c*). There was good agreement between the amplitudes of the first three peaks

Arrows were positioned at the highest point of each peak, or if two very closely spaced peaks occurred, at an intermediate point determined by their relative heights. The positions of the prominent peaks show good agreement for both pulses before and after LTP induction.

both before and after LTP, although the probabilities associated with them were redistributed, reflecting the increased quantal content for the second pulse. Thus, not only was paired pulse facilitation detected in the reduction in probability of failures, but also in the relative probabilities of the peaks making up the body of the e.p.s.c. histogram.

The relative contribution of increases in quantal content and

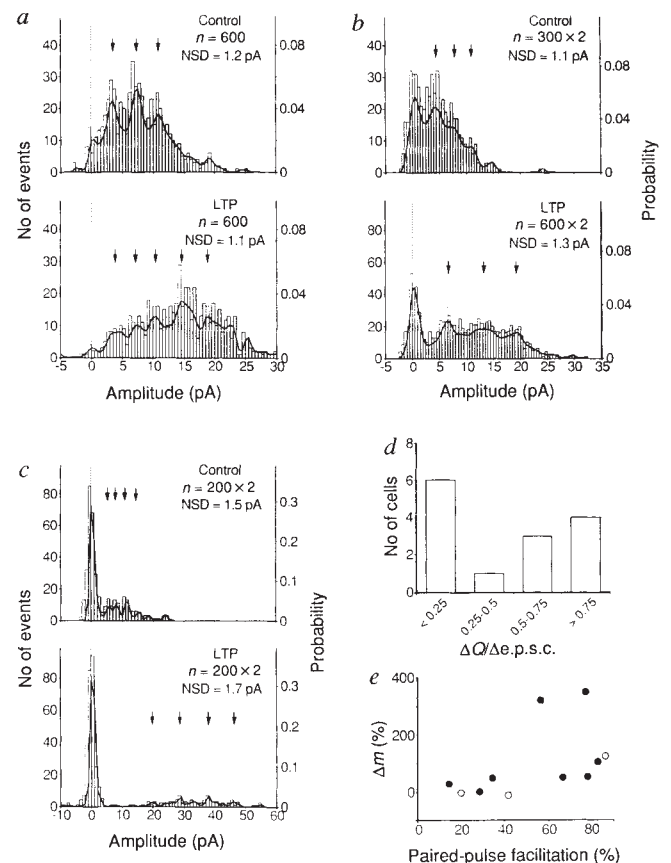


FIG. 4 Different patterns of expression of LTP. *a*–*c*, The top histograms and MEND solutions were obtained in the control period, whereas those shown below were obtained after LTP induction, with the number of sweeps and noise standard deviation indicated. *a*, Exclusive increase in quantal content, with no change in quantal amplitude. *b*, Increase in both quantal content and quantal amplitude. *c*, Exclusive increase in quantal amplitude with no change in quantal content. *d*, Bar graph plotting the number of cells against the degree to which LTP was accounted for by an increase in quantal amplitude. This was estimated by dividing the fractional change in quantal amplitude by the fractional change in the average e.p.s.c. amplitude. Results ranged from predominant increases in quantal content (left) to predominant increases in quantal amplitude (right). *e*, Per cent increase in quantal content during LTP as a function of the amount of paired-pulse facilitation in different experiments. Quantal content was estimated by dividing the average e.p.s.c. amplitude by the quantal amplitude. Some results, indicated by open symbols, were obtained in a higher Ca^{2+} concentration (4 mM instead of 2.5), which had previously been noted to reduce the average degree of paired pulse facilitation²³. Pooling the results obtained at the two Ca^{2+} concentrations, the slope of the relationship was significantly different from 0 at $P < 0.02$ (Spearman's rank correlation).

quantal amplitude to LTP varied considerably among experiments. In four cases a selective increase in quantal content occurred. An example of such a selective change is shown in Fig. 4a. In this cell there was a clear reduction in failures and a shift of the histogram to larger events, but no change in the separation of peaks. In four other cases there was a change only in quantal amplitude, with no change in quantal content (Fig. 4c). In these cases paired pulse facilitation was still associated with a decrease in failures, suggesting that if quantal content had increased with LTP it would have been detected. In the six remaining cases in which quantal peaks could be detected, both quantal content and amplitude increased after LTP (Figs 3, 4b).

A comparison of cells revealed no relationship between the increases in quantal content and quantal amplitude, implying that they represent two separate mechanisms of expression of LTP. The number of evoked responses successfully resolved is shown in Fig. 4d as a function of the degree to which the potentiation could be accounted for by changes in quantal amplitude. A positive relation was, however, observed between the degree to which quantal content increased and the paired-pulse facilitation ratio in the baseline (Fig. 4e), suggesting that the component of LTP mediated by changes in quantal content shares some mechanisms with frequency-dependent changes in transmission.

Although the possible mechanisms underlying the two forms of LTP have not been addressed in this study, our results agree with evidence obtained from several sources in favour of both pre- and postsynaptic sites of expression⁸⁻¹⁶. In particular, they indicate that an increase in quantal amplitude, as indicated by the analysis of spontaneous miniature e.p.s.cs¹⁶, does not account entirely for LTP in all cases, and that an additional increase in quantal content must be postulated. The conventional interpretation of these data is that during LTP both an increase in transmitter release (quantal content) and postsynaptic transmitter sensitivity (quantal amplitude) can occur. Alternative explanations for these results should, however, also be considered. For instance, the recruitment of patches of non-NMDA glutamate receptors could, in theory, explain an increase in quantal content and an increase in vesicle filling could explain an increase in quantal amplitude. □

The spread of Na⁺ spikes determines the pattern of dendritic Ca²⁺ entry into hippocampal neurons

David B. Jaffe*, Daniel Johnston*,
Nechama Lasser-Ross†, John E. Lisman‡,
Hiroyoshi Miyakawa§ & William N. Ross†||

* Division of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA

† Department of Physiology, New York Medical College, Valhalla, New York 10595, USA

‡ Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA

§ Department of Physiology, Yamagata University School of Medicine, 2-2 Iida Nishi, Yamagata City, Yamagata 990-23, Japan, and The Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

THE dendrites of many types of neurons contain voltage-dependent Na⁺ and Ca²⁺ conductances that generate action potentials (see ref. 1 for review). The function of these spikes is not well understood, but the Ca²⁺ entry stimulated by spikes probably affects Ca²⁺-dependent processes in dendrites. These include synaptic plasticity^{2,3}, cytotoxicity⁴ and exocytosis⁵. Several lines of evidence suggest that dendritic spikes occur within subregions of the dendrites⁶⁻⁸. To study the mechanism that govern the spread of spikes in the dendrites of hippocampal pyramidal cells, we imaged Ca²⁺ entry with Fura-2 (ref. 9) and Na⁺ entry with a newly developed Na⁺-sensitive dye¹⁰. Our results indicate that Ca²⁺ entry into dendrites is triggered by Na⁺ spikes that actively invade the dendrites. The restricted spatial distribution of Ca²⁺ entry seems to depend on the spread of Na⁺ spikes in the dendrites, rather than on a limited distribution of Ca²⁺ channels. In addition, we have observed an activity-dependent process that modulates the invasion of spikes into the dendrites and progressively restricts Ca²⁺ entry to more proximal dendritic regions.

To monitor Ca²⁺ entry, CA1 pyramidal cells were injected with Fura-2 and fluorescence was measured with a high-speed charge-coupled device camera¹¹. Figure 1a illustrates the entry of Ca²⁺ triggered by a single Na⁺ spike ($n=4$). The elevation in internal calcium ion concentration ($[Ca^{2+}]_i$) following single Na⁺ spikes occurred within less than 25 ms, the minimum frame interval in this experiment¹¹. Prolonged (0.5 s) depolarizing current pulses produced continuous trains of Na⁺ spikes and steady rises in $[Ca^{2+}]_i$ (Fig. 1b). The degree of change in $[Ca^{2+}]_i$, however, varied as a function of the distance from the soma (see later). Each spike contributed a roughly equal increment in $[Ca^{2+}]_i$, so that the cumulative Ca²⁺ rise due to spike-linked Ca²⁺ entry could be quite large, depending on the frequency of firing and the duration of the train.

The spatial distribution of spike-driven $[Ca^{2+}]_i$ elevation was highly nonuniform. For the cell shown in Fig. 1b, the largest increase in $[Ca^{2+}]_i$ occurred in the proximal dendrites about 100 μ m from the cell body while essentially no increase was detected at sites more distal than about 250 μ m, consistent with previous reports of synaptically activated calcium transients in guinea-pig pyramidal neurons¹². It has been suggested that this distribution is due to the low density of Ca²⁺ channels in distal dendrites^{12,13}. To test this possibility, we blocked K⁺ channels with 10 mM extracellular tetraethylammonium ion (TEA), a procedure that permits larger depolarizations and the generation of Ca²⁺ spikes. Under these conditions a depolarizing current

Received 28 November 1991; accepted 13 March 1992.

- Bliss, T. V. P. & Lomo, T. *J. Physiol.* **232**, 331-356 (1973).
- Bliss, T. V. P. & Lynch, M. A. in *Long-term Potentiation: From Biophysics to Behavior* (eds Landfield, P. W. & Deadwyler, S. A.) 3-72 (Liss, New York, 1988).
- Nicoll, R. A., Kauer, J. A. & Malenka, R. C. *Neuron* **1**, 97-103 (1988).
- Wigström, H., Gustafsson, B., Huang, Y.-Y. & Abraham, W. C. *Acta physiol. scand.* **126**, 317-319 (1986).
- Malinow, R. & Miller, J. P. *Nature* **320**, 529-530 (1986).
- Kelso, S. R., Ganong, A. H. & Brown, T. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5326-5330 (1986).
- Sastry, B. R., Goh, J. W. & Auyeung, A. *Science* **232**, 988-990 (1986).
- Dolphin, A. C., Errington, M. L. & Bliss, T. V. P. *Nature* **297**, 496-498 (1982).
- Voronin, L. L., Kuhn, U. & Hess, G. *Neurophysiology* **22**, 341-347 (1990).
- Malinow, R. & Tsien, R. W. *Nature* **346**, 177-180 (1990).
- Bekkers, J. M. & Stevens, C. F. *Nature* **346**, 724-729 (1990).
- Malinow, R. *Science* **252**, 722-724 (1991).
- Kauer, J. A., Malenka, R. C. & Nicoll, R. A. *Neuron* **1**, 911-917 (1988).
- Muller, D., Joly, M. & Lynch, G. *Science* **242**, 1694-1697 (1988).
- Foster, T. C. & McNaughton, B. L. *Hippocampus* **1**, 79-91 (1991).
- Manabe, T., Renner, P. & Nicoll, R. A. *Nature* **355**, 50-55 (1992).
- Edwards, F. *Nature* **350**, 271-272 (1991).
- Korn, H. & Faber, D. S. *Trends Neurosci.* **14**, 439-445 (1991).
- Clements, J. D. *Nature* **353**, 396 (1991).
- Bekkers, J. M., Richerson, G. B. & Stevens, C. F. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5359-5362 (1990).
- Larkman, A., Stratford, K. & Jack, J. *Nature* **350**, 344-347 (1991).
- Redman, S. *Physiol. Rev.* **70**, 165-198 (1990).
- Zucker, R. S. *J. Physiol.* **229**, 787-810 (1973).
- Kullmann, D. M. *J. Neurosci. Meth.* **30**, 231-245 (1989).
- Dempster, A. P., Laird, N. M. & Rubin, D. B. *J. R. statist. Soc. B* **39**, 1-21 (1977).
- Gull, S. F. & Daniell, G. *J. Nature* **272**, 686-690 (1978).
- Nicoll, R. A. & Alger, B. E. *J. Neurosci. Meth.* **4**, 153-156 (1981).
- Coleman, P. A. & Miller, R. F. *J. Neurophysiol.* **61**, 218-230 (1989).
- Blanton, M. G., Lo Turco, J. J. & Kriegstein, A. R. *J. Neurosci. Meth.* **30**, 203-210 (1989).
- Horn, R. & Marty, A. *J. gen. Physiol.* **92**, 145-159 (1988).

ACKNOWLEDGEMENTS. Supported by the NIMH and NIH (R.A.N.) and the American Epilepsy Society, Human Frontiers Science Program Organization and Smith and Nephew Foundation (D.M.K.). R.A.N. is a member of the Keck Center for Integrative Neuroscience and the Silvio Conte Center for Neuroscience Research.

|| To whom correspondence should be addressed.

step caused a mixture of Na^+ and Ca^{2+} spikes (Fig. 2a). Ca^{2+} entry during the Na^+ spikes was restricted to proximal dendrites, as seen in Fig. 1b. In contrast, Ca^{2+} entry during Ca^{2+} spikes also occurred in more distal dendrites ($n = 17$). In experiments in which the image frame was moved to explore the most distal dendrites (Fig. 2b), Ca^{2+} spikes produced Ca^{2+} entry up to $415 \pm 27 \mu\text{m}$ from the soma ($n = 4$). Because the apical dendrites of CA1 pyramidal cells from adult rat hippocampus extend about $400 \mu\text{m}$ from the soma¹⁴, Ca^{2+} channels appear to extend to the outermost apical dendrites. Other experiments showed that Ca^{2+} channels also extend into the outermost basilar dendrites. Ca^{2+} entry evoked by spikes was not reduced by $50 \mu\text{M}$ D,L-2-amino-5-phosphonovaleric acid, indicating that Ca^{2+} entry was through voltage-gated Ca^{2+} channels rather than NMDA (*N*-methyl-D-aspartate) channels activated by ambient glutamate¹⁵. These results therefore argue that voltage-gated Ca^{2+} channels are present throughout the dendrites and that the failure of Na^+ spikes to evoke Ca^{2+} entry in distal dendrites cannot be attributed to an absence of Ca^{2+} channels in distal dendrites.

An alternative explanation for the absence of distal Ca^{2+} entry in normal saline is that distal dendrites are not invaded by Na^+ spikes. To test this possibility, we monitored the spread of Na^+

spikes using SBFI (sodium-binding benzofuran isophthalate), a dye that changes its fluorescence when it binds sodium ion^{10,16}. Figure 3a shows that a single Na^+ spike caused a rapid rise in $[\text{Na}^+]_i$ in proximal dendrites. The fact that Na^+ signals were blocked by $1 \mu\text{M}$ tetrodotoxin (TTX), even during large Ca^{2+} spikes (Fig. 3b and c), indicates that the signals we detected were due to Na^+ entry through Na^+ channels. Our results thus provide direct evidence for active Na^+ spikes in pyramidal cell dendrites, which has previously been suggested on other grounds^{6,7,17-19}. The spatial spread of Na^+ spikes is shown in Fig. 3d. Na^+ entry could only be detected in the proximal dendrites, even during Ca^{2+} spikes that depolarized distal dendrites. In seven cells in which we examined the spatial distribution of $[\text{Na}^+]_i$ signals in the apical dendrites, with or without TEA in the bath, the most distal Na^+ signals averaged $208 \pm 32 \mu\text{m}$ from the cell body. These results indicate that Na^+ spikes do not actively invade distal dendrites. Because the space constant for fast Na^+ spikes is quite short in dendrites²⁰, there would be little depolarization and thus little Ca^{2+} entry expected in distal dendrites from passively propagated Na^+ spikes.

Even though Na^+ channels exist on the proximal dendrites, Na^+ spikes did not always invade these dendrites fully. During a spike train, early spikes produced a rise in $[\text{Ca}^{2+}]_i$ and $[\text{Na}^+]_i$ throughout the proximal dendrites, but later spikes produced a further rise only in regions progressively closer to the soma (Fig. 4a, b). These results indicate that there is an activity-dependent process that progressively limits the extent to which Na^+ spikes invade the proximal dendrites.

Our results show that dendritic Na^+ spikes produce large changes in $[\text{Ca}^{2+}]_i$. Moreover, results to be presented elsewhere indicate that during synaptic stimulation, voltage-gated Ca^{2+} entry contributes much more to the overall rise in dendritic $[\text{Ca}^{2+}]_i$ than entry through the NMDA channel²¹. It is therefore important to understand the mechanisms that govern this process. These experiments show that the spatial distribution of $[\text{Ca}^{2+}]_i$ accumulation is determined by the spatial distribution of Na^+ spikes rather than only on the distribution of Ca^{2+} channels. The factors that determine the spread of Na^+ spikes are complex. The distal portion of apical dendrites apparently lacks a sufficient density of Na^+ channels to support active spike invasion. In addition, the spread of Na^+ spikes is limited by an activity-dependent process. One likely candidate for this process is the Ca^{2+} -activated K^+ conductance that turns on during spike trains²².

The role of spike-driven Ca^{2+} entry is only beginning to be

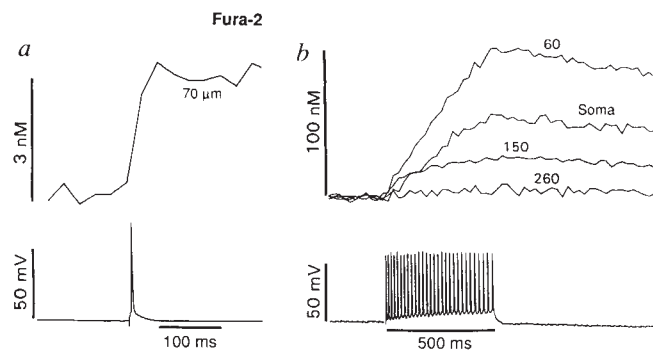


FIG. 1 Na^+ spikes trigger Ca^{2+} entry. *a*, $[\text{Ca}^{2+}]_i$ elevation $70 \mu\text{m}$ from the soma produced by a single Na^+ spike (stimulus 5 ms, 0.5 nA current pulse through an intracellular microelectrode; average of 10 given at 10-s intervals). Frame interval was 25 ms. In other experiments the rise time of the Ca^{2+} transient arising from a single spike was determined to be less than 10 ms. *b*, $[\text{Ca}^{2+}]_i$ elevation measured at different distances from the soma (in μm) caused by a train of spikes (stimulus 500 ms, 0.4 nA). At other distances the spatial profile of peak $[\text{Ca}^{2+}]_i$ was smoothly interpolated among these curves.

METHODS. Transverse hippocampal slices ($300 \mu\text{m}$) were prepared from 100–125 g male Sprague-Dawley rats²⁶ and maintained in a submerged chamber with continuous superfusion of oxygenated saline (in mM): (124) NaCl, (2.5) KCl, (26) NaHCO_3 , (2) MgCl_2 , (2) CaCl_2 , (1.25) NaH_2PO_4 and (10) dextrose. Temperature was maintained at 30°C . Microelectrode tips were filled with 200 mM potassium acetate and Fura-2 (2–6 mM) and back-filled with 4 M potassium acetate. After impalement, dye was injected with steady hyperpolarizing current (approximately -1 nA) for 10 to 20 min. During this time the dye diffused throughout the dendrites. Recordings were made with an Axoclamp 2A amplifier in bridge mode. High-speed images (10-ms or 25-ms frame intervals) of the changing Fura-2 fluorescence levels were recorded using a cooled CCD camera (Photometrics) operated in a sequential frame transfer mode¹¹. Changes in $[\text{Ca}^{2+}]_i$ were recorded as changes in $\Delta F/F$, where F is the fluorescence at 380 nm excitation (emission at wavelengths greater than 495 nm) at resting potential (corrected for background autofluorescence) and ΔF is the reduction from this value during neuronal activity corrected for bleaching. The $\Delta F/F$ signals for Fura-2 were converted to Ca^{2+} concentration by the formula^{9,27}

$$[\text{Ca}^{2+}]_i = \frac{[\text{Ca}^{2+}]_{\text{rest}} + K_d(\Delta F/F)/(\Delta F/F)_{\text{max}}}{(1 - (\Delta F/F)/(\Delta F/F)_{\text{max}})}$$

where $[\text{Ca}^{2+}]_{\text{rest}}$ was assumed to be 60 nM (taken from ref. 12 and our own measurements), $(\Delta F/F)_{\text{max}} = 0.8$, and $K_d = 230 \text{ nM}$. The optical traces in the figures were obtained by averaging the values for each pixel within a rectangular region of $\sim 30 \times 30 \mu\text{m}^2$ in the dendrites and $10 \times 10 \mu\text{m}^2$ in the soma (see ref. 11).

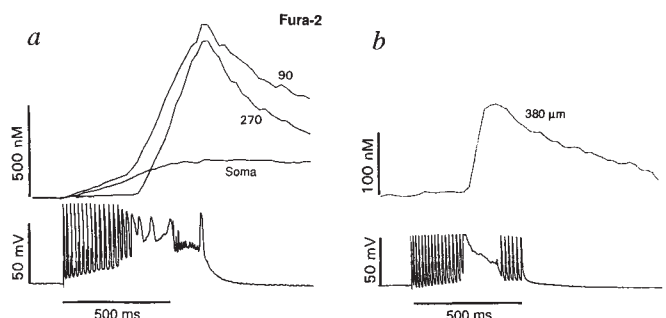
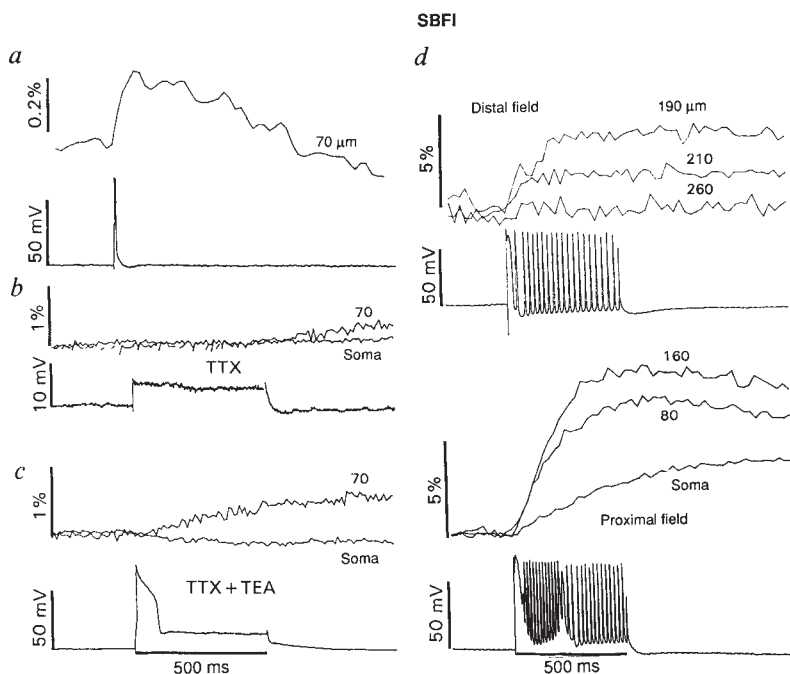


FIG. 2 Spatial distribution of dendritic Ca^{2+} entry produced by Na^+ and Ca^{2+} spikes. *a*, Ca^{2+} entry after application of TEA to facilitate the occurrence of Ca^{2+} spikes. A mixture of short-duration Na^+ spikes and long-duration Ca^{2+} spikes were evoked by a 500 ms, 0.4 nA depolarizing current pulse after bath application of 10 mM TEA and $1 \mu\text{M}$ CNQX (6-cyano-7-nitroquinoxaline-2,3-dione; to prevent epileptiform discharges). Spike-driven Ca^{2+} entry is shown at 3 different locations in the cell (soma and 90 and $270 \mu\text{m}$ from the soma). *b*, Ca^{2+} entry into very distal dendrites ($380 \mu\text{m}$ from soma) during Na^+ and Ca^{2+} spikes. A 500 ms, 0.3 nA pulse was given after adding 10 mM TEA and $1 \mu\text{M}$ CNQX. (Same cell as used for Fig. 1b.)

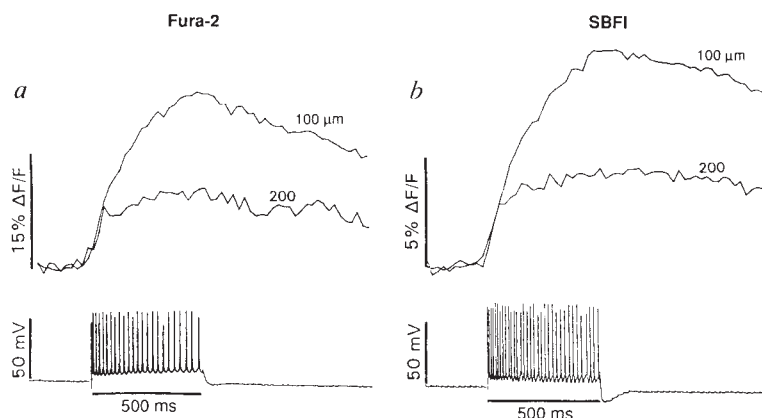
FIG. 3 Na^+ entry measured with SBFI. *a*, Na^+ entry in proximal apical dendrites ($70\ \mu\text{m}$ from the soma) following a single Na^+ spike. The trace is an average of 18 spikes (stimulus $10\ \text{ms}$, $0.8\ \text{nA}$) stimulated every $5\ \text{s}$. *b*, TTX blocks Na^+ entry. Average of 10 depolarizing current steps ($500\ \text{ms}$, $1\ \text{nA}$) in the presence of $1\ \mu\text{M}$ TTX. Na^+ entry from the soma and from a dendritic region $70\ \mu\text{m}$ from the soma is illustrated. Figure 4*b* shows large signals from the same cell before TTX was added. *c*, TEA ($25\ \text{mM}$) was then added to the bath and the changes in fluorescence averaged for 10 current pulses ($1.4\ \text{nA}$, $500\ \text{ms}$ duration) given every $30\ \text{s}$. Large Ca^{2+} spikes were triggered by the current pulses, but no Na^+ entry was detected in the soma. The small delayed rise in dendritic $[\text{Na}^+]_i$ may be due to $\text{Na}^+/\text{Ca}^{2+}$ exchange activated by the Ca^{2+} entry during the Ca^{2+} spike²⁸. *d*, Spatial distribution of Na^+ entry. An average of 5 current pulses ($500\ \text{ms}$, $0.4\ \text{nA}$, each of which triggered a combination of Na^+ and Ca^{2+} spikes), was given to a CA1 neuron with the slice bathed in $10\ \text{mM}$ TEA and $1\ \mu\text{M}$ CNQX. The $\Delta F/F$ signals at 3 different locations (soma and 80 and $160\ \mu\text{m}$ from the soma) are illustrated for the 'proximal field'. A similar average of 5 depolarizing current pulses was given to the same cell after moving the image frame to more distal regions of the apical dendrites ('distal field'). The $\Delta F/F$ signals at 190 , 210 and $260\ \mu\text{m}$ from the soma are illustrated. For both sets of records the temporal configuration of the Na^+ and Ca^{2+} spikes was almost identical in each of the five trials. It is unlikely that $\text{Na}^+/\text{Ca}^{2+}$ exchange contributes substantially to the entry of Na^+ during a 500-ms train of spikes, because results similar to those shown here were obtained in other experiments with zero Ca^{2+} added in the saline.

METHODS. Electrode tips were filled with $12\ \text{mM}$ SBFI and $200\ \text{mM}$ potassium acetate. Cells were filled with dye by ionophoresis. In most cases SBFI fluorescence was visible out to the outermost tips of the apical dendrites. Resting fluorescence (F) and high-speed image sequences of the changing fluorescence were recorded using the same filter set as with Fura-2. Similar



corrections for background fluorescence and dye bleaching were applied. No attempt was made to calibrate $\Delta F/F$ in terms of $[\text{Na}^+]_i$ because of the unknown intracellular environment. Based on spectral data¹⁰, however, $\Delta F/F$ is monotonically related to $\Delta[\text{Na}^+]_i$.

FIG. 4 Activity-dependent changes in the location of Na^+ and Ca^{2+} entry. *a*, Response of a CA1 neuron injected with Fura-2 to a train of spikes. Ca^{2+} entry was measured in dendrites 100 and $200\ \mu\text{m}$ from the soma. At $200\ \mu\text{m}$ only the first few spikes in the train produced a detectable rise in $[\text{Ca}^{2+}]_i$, whereas at $100\ \mu\text{m}$ the whole train produced a rise in $[\text{Ca}^{2+}]_i$. Stimulus was $500\ \text{ms}$, $0.4\ \text{nA}$. Ordinate is in units of $\Delta F/F$ to facilitate comparison with the SBFI signals. *b*, Similar experiment measuring Na^+ entry with SBFI. The $\Delta F/F$ signals measured in dendrites 100 and $200\ \mu\text{m}$ from the soma (average of 10 trains). At $200\ \mu\text{m}$ only the first spikes in the train produce detectable Na^+ entry, but Na^+ entry continued during the train at more proximal locations. Stimulus was $500\ \text{ms}$, $0.5\ \text{nA}$.



investigated, so the implications of the spatial distribution of $[\text{Ca}^{2+}]_i$ we have measured are a matter of conjecture²³. One relevant fact is that different types of synaptic inputs are located at different distances from the soma. For instance, perforant path synapses onto CA1 pyramidal cells in rats are over $300\ \mu\text{m}$ from the soma²⁴ where spike-driven $[\text{Ca}^{2+}]_i$ accumulation is negligible, whereas Schaffer collateral inputs are closer to the soma²⁵ where the Ca^{2+} influx is large. These facts suggest that the location of synapses along the proximal-distal axis is likely to have important functional consequences. □

Received 2 December 1991; accepted 25 March 1992.

1. Llinas, R. R. *Science* **242**, 1654-1664 (1988).
2. Malenka, R. C., Kauer, J. A., Zucker, R. S. & Nicoll, R. A. *Science* **242**, 81-84 (1988).
3. Lynch, G., Larson, J., Kelso, S., Barrionuevo, G. & Schottler, F. *Nature* **305**, 719-721 (1983).
4. Choi, D. W. *Trends Neurosci.* **11**, 465-469 (1988).
5. Llinas, R. & Greenfield, S. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3047-3050 (1987).
6. Miyakawa, H. & Kato, H. *Brain Res.* **399**, 303-309 (1986).
7. Turner, R. W., Meyers, D. E. R. & Barker, J. L. *J. Neurophysiol.* **62**, 1375-1387 (1989).
8. Ross, W. N., Lasser-Ross, N. & Werman, R. *Proc. R. Soc.* **240**, 173-185 (1990).

9. Grynkiewicz, G., Poenie, M. & Tsien, R. Y. *J. biol. Chem.* **260**, 3440-3450 (1985).
10. Minta, A. & Tsien, R. Y. *J. biol. Chem.* **264**, 19449-19457 (1989).
11. Lasser-Ross, N., Miyakawa, H., Lev-Ram, V., Young, S. R. & Ross, W. N. *J. neurosci. Meth.* **36**, 253-261 (1991).
12. Regehr, W. G., Connor, J. A. & Tank, D. W. *Nature* **341**, 533-536 (1989).
13. Westenbroek, R. E., Ahlman, M. K. & Catterall, W. A. *Nature* **347**, 281-284 (1990).
14. Turner, R. W., Meyers, D. E. R., Richardson, T. L. & Barker, J. L. *J. Neurosci.* **11**, 2270-2280 (1991).
15. Sah, P., Hestrin, S. & Nicoll, R. A. *Science* **246**, 815-818 (1989).
16. Lasser-Ross, N. & Ross, W. N. *Proc. R. Soc.* **247**, 35-39 (1992).
17. Andersen, P., Storm, J. & Wheal, H. V. *J. Physiol., Lond.* **383**, 509-526 (1987).
18. Wong, R. K. S. & Prince, D. A. *Science* **204**, 1228-1231 (1979).
19. Masukawa, I. M. & Prince, D. A. *J. Neurosci.* **4**, 217-227 (1984).
20. Johnston, D. & Brown, T. H. *J. Neurophysiol.* **50**, 464-486 (1983).
21. Ross, W. N., Jaffe, D., Lasser-Ross, N., Lisman, J. & Johnston, D. *Soc. Neurosci. Abstr.* **17**, 581 (1991).
22. Lancaster, B. & Adams, P. R. *J. Neurophysiol.* **55**, 1268-1282 (1986).
23. Lisman, J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9574-9578 (1989).
24. Steward, O. *J. comp. Neurol.* **167**, 285-314 (1976).
25. Amaral, D. G. & Witter, M. P. *Neuroscience* **31**, 571-591 (1989).
26. Jaffe, D. & Johnston, D. *J. Neurophysiol.* **64**, 948-960 (1990).
27. Lev-Ram, V., Miyakawa, H., Lasser-Ross, N. & Ross, W. N. *Soc. Neurosci. Abstr.* **16**, 894 (1990).
28. Blaustein, M. P. *Trends Neurosci.* **11**, 438-443 (1988).

ACKNOWLEDGEMENTS. We thank I. Levitan and E. Marder for critically reading the manuscript. This work was supported by grants from the NIH and NIMH.

An immunological role for the CD8 β -chain

Christopher J. Wheeler, Paul von Hoegen*
& Jane R. Parnes†

Department of Medicine, Division of Immunology and Rheumatology,
Stanford University Medical Center, Stanford, California 94305, USA

MATURE T cells can be functionally divided into two categories distinguished by surface expression of either CD4 or CD8, which in turn corresponds to restriction by and binding to class II or class I major histocompatibility complex proteins, respectively¹⁻⁸. CD8 can be expressed as a homodimer of the α -chain, or as a heterodimer of α - and β -chains on human and mouse T cells, although most peripheral T cells seem to express CD8 $\alpha\beta$ heterodimers exclusively (reviewed in ref. 9). Functional characterization of CD8 has focused primarily on the effect of the α -chain, which enhances or reconstitutes T-cell responses in homodimeric form^{10,11} and may play a specific role in thymic selection¹²⁻¹⁴. In contrast, no role has been ascribed to CD8 β or $\alpha\beta$ heterodimers specifically. Here we show that CD8 $\alpha\beta$ transfectants produce more interleukin-2 than CD8 α transfectants in response to specific stimuli. Increased interleukin-2 is also observed in cells expressing hybrid CD8 β - α molecules (extracellular CD8 β plus CD8 α transmembrane and cytoplasmic regions) on their surface. These results indicate that external portions of CD8 β could be critical and that they may act independently of CD8 α in mediating their augmentation effect.

The allogeneic CD4⁺, CD8⁻ mouse T-cell hybridoma HTB-157.7 has been reported to produce interleukin-2 (IL-2) when stimulated by L cells stably transfected with H-2K^b (L-K^b), about 10% of this amount of IL-2 when stimulated by L cells transfected with H-2K^bbm10 (L-bm10), and no IL-2 in response to either H-2K^bbm1 transfectants (L-bm1) or to untransfected Ltk⁻ cells¹⁵. In our hands, however, HTB-157.7 typically produces IL-2 only in response to L-K^b, with no response to L cells expressing comparable levels of either H-2K^bbm10 or H-2K^bbm1, or to untransfected Ltk⁻ cells (Fig. 1, and data not shown). To examine the role of CD8 β in class I-restricted T-cell responses, we initially transfected HTB-157.7 with expression vectors containing either the CD8 α or CD8 β complementary DNAs separately or together in cotransfections, and derived transfectants stably expressing similar levels of CD8 α and CD3 on their surfaces, as determined by cytofluorometry (Table 1; Fig. 2a and b). Two groups of transfectants expressing equivalent amounts of both CD8 α and CD3, and which produced similar levels of IL-2 on stimulation with maximal levels of plate-bound anti-CD3 monoclonal antibody, were selected for further analysis (see legend to Fig. 1). When these matched CD8 α and CD8 $\alpha\beta$ transfectants were analysed for IL-2 production, we found that all gave significantly increased amounts of IL-2 in response to L-K^b (Fig. 1a), as well as a novel response to L-bm10 (Fig. 1b). But the L-K^b responses mediated by CD8 $\alpha\beta$ transfectants differed from those of CD8 α transfectants in that significantly more IL-2 was produced by the former at lower densities of stimulator cells. Secreted IL-2 levels at the highest L-K^b stimulator densities were not significantly different between CD8 α and CD8 $\alpha\beta$ transfectants. Differences between CD8 $\alpha\beta$ and CD8 α transfectants were even more striking in response to L-bm10, with significantly more IL-2 being produced by CD8 $\alpha\beta$ transfectants at most stimulator densities (Fig. 1b-f). Moreover, sorted CD8 α transfectant variants expressing as much as four times the level of CD8 α , but the same amount of

TABLE 1 Quantitation of expression of CD8 α , CD8 β and CD3 on transfected HTB-157.7

Cells	CD8 α	CD8 β	CD3
HTB-157.7	2.3	1.4	13.0
α 4	176.9	1.7	8.0
2 α 1-11	221.4	2.0	9.3
$\alpha\beta$ 7	210.6	17.8	7.8
2 α 1-12	149.7	1.8	9.5
$\alpha\beta$ 1	147.2	29.6	7.8
β - α 1	2.2	10.1	8.8
β - α 2-4	2.0	4.9	7.7
β - α 2-5	2.5	4.7	13.2

Cells were stained as described in Fig. 2 legend. Data are presented as mean fluorescence intensity at equivalent gain. Matched CD8 α and CD8 $\alpha\beta$ transfectants were stained simultaneously. Surface levels of CD3, CD8 α and/or CD8 β were somewhat variable, and matched groups were chosen on the basis of multiple analyses. From these analyses we can estimate the level of CD8 β epitopes expressed on CD8 β - α hybrid transfectants to be about twofold to threefold lower than observed on either of the CD8 $\alpha\beta$ transfectants.

CD3 as their CD8 $\alpha\beta$ counterparts, gave response patterns indistinguishable from those of the original CD8 α transfectants (data not shown). This ensures that any enhanced response in CD8 $\alpha\beta$ transfectants compared with CD8 α transfectants is not due simply to the presence of more CD8 complexes on the former.

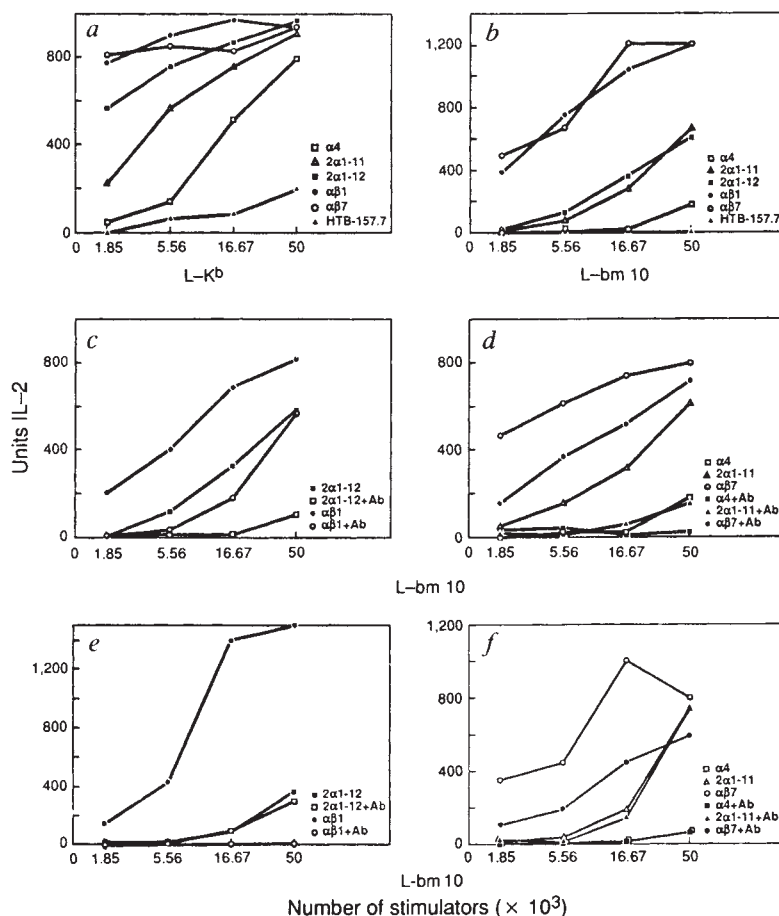
Antibody blocking studies revealed another difference between CD8 α and CD8 $\alpha\beta$ transfectants. Although addition of anti-CD8 α monoclonal antibody during stimulation inhibited IL-2 responses of all transfectants to L-bm10, inhibition of CD8 $\alpha\beta$ transfectants was typically incomplete, whereas inhibition of CD8 α transfectants was complete or nearly so at identical concentrations of antibody (Fig. 1c and d). Inhibition of transfectant variants expressing increased levels of CD8 α by anti-CD8 α monoclonal antibody did not differ significantly with respect to their parental transfectants, indicating that this antibody was not limiting. In contrast, a range of inhibition by monoclonal antibody against CD8 β was observed on CD8 $\alpha\beta$ transfectants in response to L-bm10. Inhibition, when incomplete, was most dramatic at lower densities of L-bm10, and in general correlated inversely with surface CD8 α levels on CD8 $\alpha\beta$ transfectants (Fig. 1e and f; and data not shown). As expected, CD8 α transfectants were not inhibited by the antibody against CD8 β . Neither anti-CD8 α nor anti-CD8 β antibodies inhibited the basal IL-2 response of HTB-157.7 to L-K^b stimulators (data not shown).

We next investigated whether the extracellular portions of CD8 β could mediate augmentative effects on HTB-157.7 independently of those of CD8 α . As CD8 β is not normally expressed on cell surfaces in the absence of CD8 α (ref. 16), we created a DNA construct, p β RV α , that encodes the external portions of CD8 β together with the transmembrane and cytoplasmic portions of CD8 α , in an attempt to obtain only CD8 β expression on the cell surface. Transfection of p β RV α led to specific expression of CD8 β epitopes in the complete absence of CD8 α on the surface of HTB-157.7 (Fig. 2c). Transfectants produced consistently elevated levels of IL-2 only in response to L-K^b (Fig. 3). The level of IL-2 produced by p β RV α transfectants was roughly equivalent to that produced by transfectant α 4 in response to L-K^b. But unlike α 4 and the other transfectants, p β RV α transfectants produced no IL-2 in response to L-bm10. Thus, the response mediated by the β - α hybrid molecules is both different from and independent of that mediated by CD8 α or CD8 $\alpha\beta$. As functional augmentation by CD8 α requires interaction of extracellular portions with major histocompatibility complex (MHC) class I ligands⁶⁻⁸, as well as signalling through cytoplasmic portions¹⁷⁻¹⁹, the fact that we see similar augmentation with our β - α hybrid transfectants

* Present address: Institut für Immunologie und Genetik 07/710, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D/6900 Heidelberg, Germany.
† To whom correspondence should be addressed.

FIG. 1 IL-2 responses of CD8 α and CD8 $\alpha\beta$ transfectants. Groups of CD8 α and CD8 $\alpha\beta$ transfectants matched for CD8 α and CD3 surface expression were analysed on *a*, L-K^b and *b*, L-bm10 stimulators a minimum of 9 times: $\alpha 4$ and 2 $\alpha 1$ -11 were compared with $\alpha\beta 7$, and 2 $\alpha 1$ -12 was compared with $\alpha\beta 1$. IL-2 levels generally correlated directly with stimulator cell density. Similarly, stimulation by L-K^b sublines that differed in surface levels of H-2K^b resulted in response levels which directly correlated with surface alloantigen levels (data not shown). Results similar to those presented were obtained with an additional group comprising six CD8 $\alpha\beta$ and two CD8 α transfectants. Within the range of CD8 α and CD8 β expression observed on our CD8 $\alpha\beta$ transfectants, enhanced function correlates to the presence of detectable CD8 β , but not to the CD8 α :CD8 β ratio. Maximal IL-2 production on stimulation by plate-bound anti-CD3 antibody resulted in values that were not consistently different among parental HTB-157.7, CD8 α , CD8 $\alpha\beta$ and CD8 β - α hybrid transfectants. Thus, differences between CD8 α and CD8 $\alpha\beta$ transfectants were not due to defects or discrepancies in TCR-mediated signalling or IL-2 production. Blocking by monoclonal antibodies of matched sets of CD8 α and CD8 $\alpha\beta$ transfectants: *c*, 2 $\alpha 1$ -12 and $\alpha\beta 1$ plus anti-CD8 α antibody 53-6.72 (ref. 26). *d*, $\alpha 4$, 2 $\alpha 1$ -11 and $\alpha\beta 7$ plus anti-CD8 α antibody 53-6.72 (ref. 26). *e*, 2 $\alpha 1$ -12 and $\alpha\beta 1$ plus anti-CD8 β antibody YTS-156.7.7 (H. Waldmann, unpublished results). *f*, $\alpha 4$, 2 $\alpha 1$ -11 and $\alpha\beta 7$ plus anti-CD8 β antibody YTS-156.7.7. Similar blocking results were obtained with transfectant 3 $\alpha\beta 2$ (data not shown). Relative differences between IL-2 responses of CD8 α and CD8 $\alpha\beta$ transfectants, as illustrated by these representative assays, were observed in every assay.

METHODS. DNA construction and transfection: the pBRV α construct encoding the hybrid CD8 β - α molecule was engineered by creating EcoRV sites at the extracellular/transmembrane junctions of the mouse CD8 α and CD8 β cDNAs and excising and replacing the 3' CD8 β EcoRV-SalI (vector) fragment with that from CD8 α . cDNAs encoding mouse CD8 α and CD8 β were previously subcloned¹⁹ into the expression vector pHBAPr-2-neo²⁷. DNA (50 μ g) was transfected into 8×10^6 HTB-157.7 cells by electroporation, and transfectants were selected as described¹⁹. Functional analysis of HTB-157.7 and IL-2 assay: alloantigen expression on stimulator L-cell transfectants was confirmed with monoclonal antibodies 20-8-4 (anti-H-2K^b, H-2K^b^{bm10}, H-2K^b^{bm11}) (ref. 28) and 128-9-10 (anti-H-2K^b^{bm10}) (ref. 29). Stimulator cells were irradiated with 4,500 rads, plated into microtitre wells in 100 μ l RPMI medium at the indicated densities and allowed to settle for 3-18 h at 37 °C. HTB-157.7 cells (100 μ l untransfected and transfected) at 10^6 cells per ml were added to triplicate wells and incubated at 37 °C, 18-24 h, at which time 100 μ l supernatant was



removed for IL-2 quantitation using the IL-2-dependent line HT-2, as previously described^{19,30}. Purified monoclonal antibody used for blocking studies was added to wells before introduction of HTB-157.7 cells at concentrations of 0.5 or 1.86 μ g ml⁻¹ (53-6.72) and 16 μ g ml⁻¹ (YTS-156.7.7). Stimulation by plate-bound anti-CD3 antibody: HTB-157.7 cells (100 μ l) at 10^6 cells per ml were added to wells previously coated with 10 μ g ml⁻¹ 145-2C11 or γ CD3.1 (refs 31, 32) (1.5 h at 37 °C), and incubated at 37 °C, 5% CO₂ 18-24 h. Supernatant (100 μ l) was removed for IL-2 quantitation as described.

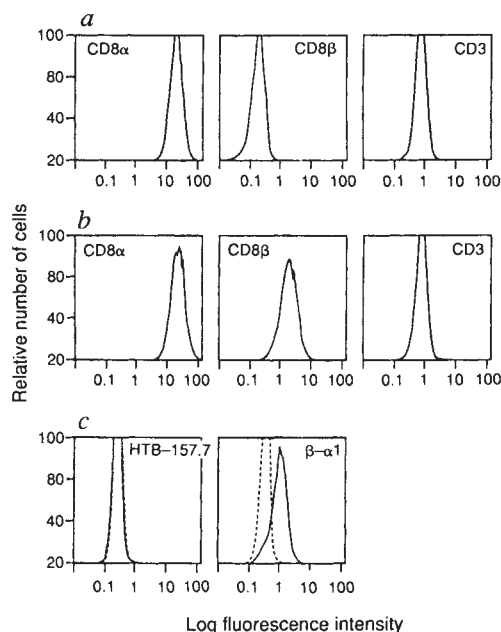


FIG. 2 Surface expression of CD8 α , CD8 β and CD3 on transfected HTB-157.7. HTB-157.7 cells stably transfected with CD8 α ($\alpha 4$), or with both CD8 α and CD8 β ($\alpha\beta 7$), were analysed for surface expression of CD8 α (monoclonal antibody 2.43; ref. 33), CD8 β (antibody 53-5.8 (ref. 26) and YTS-156.7.7 (H. Waldmann, unpublished)) and CD3 (antibody 145-2C11; ref. 32). Transfectants $\alpha 4$ (*a*) and $\alpha\beta 7$ (*b*) are representative of CD8 α - and CD3-matched transfectants. HTB-157.7 cells stably transfected with pBRV α were analysed for surface expression with a variety of anti-CD8 β antibodies. (*c*) Specific staining was obtained with YTS-156.7.7. Primary antibody was incubated with cells on ice for 15-45 min, followed by washes with PBS plus 10% fetal calf serum, secondary staining with fluorescein-conjugated mouse-anti-rat IgG, washing, and analysis by cytofluorometry. Transfectants were analysed at least 4 times with similar results. Background staining (cells plus secondary antibody) is depicted by dotted lines. HTB-157.7 is the untransfected hybridoma; β - α 1 is a representative pBRV α transfectant.

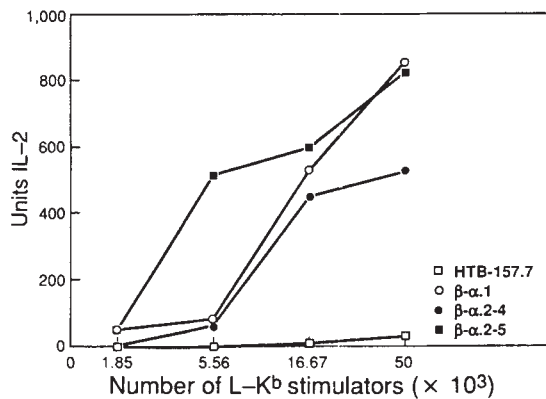


FIG. 3 IL-2 responses of hybrid CD8 β - α transfectants. Transfectants were analysed on L-K^b and L-bm10 stimulators as described in the legend to Fig. 2: β - α .1, β - α .2-4 and β - α .2-5 expressed similar levels of surface CD3 (monoclonal antibody yCD3.1) and produced significantly increased levels of IL-2 in response to L-K^b stimulators only. Each transfectant was assayed a minimum of 3 times: For β - α .1, $n=8$; for β - α .2-4, $n=5$; and for β - α .2-5, $n=3$. An additional transfectant, β - α .2-2, gave similar results in a single assay, but subsequently lost CD3 expression before storage. A representative assay is shown.

enables us to conclude that the extracellular portions of CD8 β mediate a critical interaction resulting in increased IL-2 production, presumably by signalling through the cytoplasmic portions of CD8 α . A similar effect mediated by extracellular portions of CD8 β is probably responsible for the augmented IL-2 response in CD8 $\alpha\beta$ transfectants.

In conclusion, we have demonstrated a significant difference in the responsiveness of CD8 $\alpha\beta$ versus CD8 α transfectants to specific T-cell antigen receptor (TCR)-mediated responses: cell-surface expression of CD8 β gives increased IL-2 production when stimulated by alloantigen-expressing L-cell stimulators. This difference is most striking under conditions in which the TCR-MHC interaction is most likely to be limiting, as demonstrated by IL-2 responses at low densities of L-K^b stimulators, as well as responses to L-bm10 stimulators. (H-2K^b differs from the original alloantigen H-2K^b by five amino-acid substitutions localized within the MHC-antigen binding cleft with which the allo-TCR on HTB-157.7 is likely to interact^{20,21}.) Moreover, enhanced IL-2 production mediated by surface expression of CD8 β - α hybrid molecules on HTB-157.7 suggests that extracellular portions of CD8 β act independently of those of CD8 α to augment TCR-mediated responses, as reflected here by IL-2 production. The augmentation afforded by external CD8 β epitopes in this system may proceed by either or both of two mechanisms that are not mutually exclusive. First, the decreased effectiveness of blocking by anti-CD8 α antibody on CD8 $\alpha\beta$ transfectants in particular may indicate that these have a higher-avidity interaction with stimulator cells than do their CD8 α counterparts, pointing to an additional adhesive role for CD8 β . By analogy with CD8 α , for example, CD8 β might interact with some portion of MHC class I molecules⁶⁻⁸. But surface expression of β - α hybrid molecules, while enhancing IL-2 production in response to the original alloantigen on L-K^b, does not afford response to L-bm10. Therefore, any independent adhesive effects provided by CD8 β epitopes, insofar as they are accurately reflected by our β - α hybrid molecule, must be less efficient than those demonstrated previously for CD8 α . Second, decreased blocking efficiency on CD8 $\alpha\beta$ transfectants may reflect lower response thresholds with respect to those of CD8 α transfectants. A CD8 β -mediated threshold effect might be achieved through increased interaction between CD8 and TCR complexes, and/or increased proximity of CD8-associated signalling molecules, such as the tyrosine kinase, p56^{lck}, on stimulation through the TCR (refs 22-25). Further analysis will be

required to determine definitively the mechanisms operative in CD8 β -mediated T-cell response enhancement. □

Received 4 November 1991; accepted 19 March 1992.

- Swain, S. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7101-7105 (1981).
- Swain, S. L. *Immunol. Rev.* **74**, 129-142 (1983).
- Dialynas, D. P. et al. *Immunol. Rev.* **74**, 29-56 (1983).
- MacDonald, H. R., Glasebrook, A. L., Bron, C., Kelson, A. & Cerottini, J.-C. *Immunol. Rev.* **68**, 89-115 (1982).
- Normant, A. & Littman, D. R. *EMBO J.* **7**, 3433-3439 (1988).
- Normant, A., Salter, R. D., Parham, P., Englehard, V. H. & Littman, D. R. *Nature* **336**, 79-81 (1988).
- Salter, R. D. et al. *Nature* **338**, 345-347 (1989).
- Salter, R. A. D. et al. *Nature* **345**, 41-46 (1990).
- Parnes, J. R. *Adv. Immun.* **44**, 265-311 (1989).
- Dembic, Z. et al. *Nature* **326**, 510-511 (1987).
- Gabert, J. et al. *Cell* **50**, 545-554 (1987).
- Sha, W. C. et al. *Nature* **335**, 271-274 (1988).
- Robey, E. A. B. et al. *Cell* **64**, 99-107 (1991).
- Fung-Leung, W.-P. et al. *Cell* **65**, 443-449 (1991).
- Schneck, J., Maloy, W. L., Coligan, J. E. & Margulies, D. H. *Cell* **56**, 47-55 (1989).
- Gorman, S. D., Sun, Y. H., Zamojska, R. & Parnes, J. R. *J. Immun.* **140**, 3646-3653 (1988).
- Zamojska, R. et al. *Nature* **342**, 278-281 (1989).
- Letourneur, R. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2339-2343 (1990).
- Miceli, M. C., von Hoegen, P. & Parnes, J. R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2623-2627 (1991).
- Geliebter, J. & Nathanson, S. G. *Molec. cell. Biol.* **8**, 4342-4352 (1988).
- Bjorkman, P. J. et al. *Nature* **239**, 512-518 (1987).
- Veillelette, A., Bookman, M. B., Horak, E. M. & Bolen, J. B. *Cell* **55**, 301-308 (1988).
- Veillelette, A., Bookman, M. B., Horak, E. M., Samelson, L. E. & Bolen, J. B. *Nature* **338**, 257-259 (1989).
- Barber, E. K., Dasgupta, J. D., Schlossman, S. F., Trevillyan, J. M. & Rudd, C. E. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3277-3281 (1989).
- Luo, K. & Sefton, B. M. *Molec. cell. Biol.* **10**, 5305-5313 (1990).
- Ledbetter, J. A. & Herzenberg, L. A. *Immunol. Rev.* **46**, 63-90 (1979).
- Gunning, P., Leavitt, J., Muscat, G., Ng, Y. S. & Kedes, L. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4831-4837 (1987).
- Bluestone, J. A. et al. *J. Immunogenet.* **11**, 197-207 (1984).
- Lewis, J. et al. *J. Immun.* **191**, 728-735 (1988).
- Mossman, T. R. *J. Immunol. Meth.* **65**, 55-58 (1983).
- Leo, O., Foo, M., Sachs, D. H., Samelson, L. E. & Bluestone, J. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1374-1378 (1987).
- Portoles, P. et al. *J. Immun.* **142**, 4169-4175 (1989).
- Sarmiento, M., Glasebrook, A. L. & Fitch, F. W. *J. Immun.* **125**, 2665-2672 (1984).

ACKNOWLEDGEMENTS. We thank J. Schneck for H-2 mutant-transfected L cells, R. Flavell for H-2K^b-transfected L cells, J. Bluestone, D. Sachs and H. Waldmann for monoclonal antibodies, R. Benjamin and M. Tütt-Landolfi for purified monoclonal antibodies, M. C. Miceli for technical advice and critically reading the manuscript, and T. Knaak for FACS operation. This work was supported by grants from the NIH. C.J.W. was supported by Stanford University programs in Cancer Biology and in Immunology and is a fellow of the Leukemia Society of America. J.R.P. is an Established Investigator of the American Heart Association.

Irreversible association of peptides with class II MHC molecules in living cells

Antonio Lanzavecchia*, Pamela A. Reid† & Colin Watts†

* Basel Institute for Immunology, Grenzacherstrasse 487, Postfach CH-4005, Basel, Switzerland

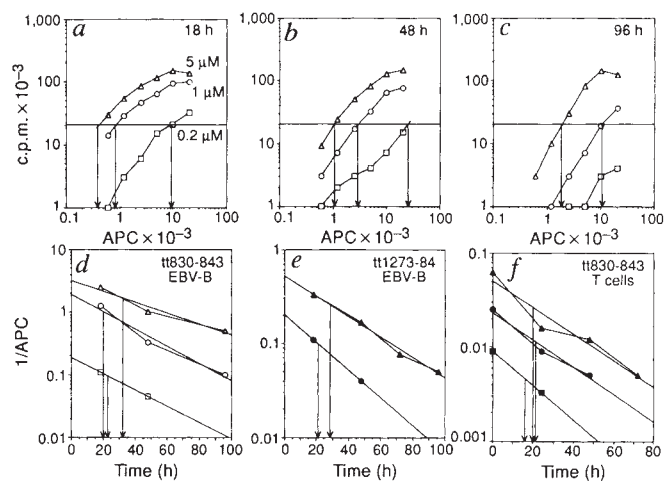
† Department of Biochemistry, Medical Sciences Institute, University of Dundee, Dundee DD1 4HN, UK

FUNCTIONAL, morphological and biochemical evidence indicates that class II major histocompatibility complex (MHC) molecules associate with processed peptides during biosynthesis¹⁻⁹. Peptide/MHC complexes in living cells have been reported to be less stable than similar complexes generated *in vitro*, which has led to the suggestion that there may be a peptide exchange mechanism operating *in vivo*¹⁰⁻¹². Although this could increase the capacity for binding incoming antigens, it would reduce the efficacy of processed antigenic peptides by exchanging these for self peptides. Here we measure the half-life of peptide/class II complexes in human antigen-presenting cells and find that it is very similar to the half-life of class II molecules themselves, indicating that peptides are bound irreversibly under physiological conditions. Thus class II MHC retains long-term 'memory' of past encounters with antigen to maximize the opportunity for T cell/antigen-presenting cell interaction.

FIG. 1 Lifetime of T-cell epitopes on living APC. *a-c*, Proliferative response of a tt830-843-specific DRw11-restricted T-cell clone to graded numbers of DRw11⁺EBV-B cells pulsed for 2 h at 37 °C with 5 μ M (Δ), 1 μ M ($\circ$) and 0.2 μ M ($\square$) tt830-843, washed and cultured in the absence of peptide for 18, 48 and 96 h. *d*, Half-life of the tt830-843-DRw11 complexes. *e*, Half-life of tt1273-84/DRw52a complexes on EBV-B cells incubated with 5 μ M ($\blacktriangle$) or 1 μ M ($\bullet$) peptide. *f*, Half-life of tt830-843/DRw11 complexes measured on activated DRw11⁺T cells that had been pulsed at 0 °C with 20 μ M ($\blacktriangle$), 5 μ M ($\bullet$) or 1.2 μ M ($\blacksquare$) peptide.

METHODS. T-cell clones and EBV-B cells were isolated and maintained as previously described³⁷. APC were pulsed with peptide in RPMI 1640, 10% FCS medium, either at 37 °C or at 0 °C and washed four times. Although the efficacy of cells pulsed at 0 °C was much lower, the half-life results were always comparable. T cells (2×10^4) were cultured with graded numbers of irradiated APC in 200 μ l RPMI, 10% FCS in 96-well flat-bottom microplates. After 48 h, the cultures were pulsed with 1 μ Ci ³H-thymidine and the radioactivity incorporated was measured after an additional 16 h by liquid scintillation. As irradiated APC we used either EBV-B cells (6,000 R), PBMC (3,000 R) or activated T cells (1,500 R). Note that T cells were taken at a time after restimulation when proliferation had virtually ceased; similarly, after washing of pulsed EBV-B cells, cell division slows to once every 2-3 days.

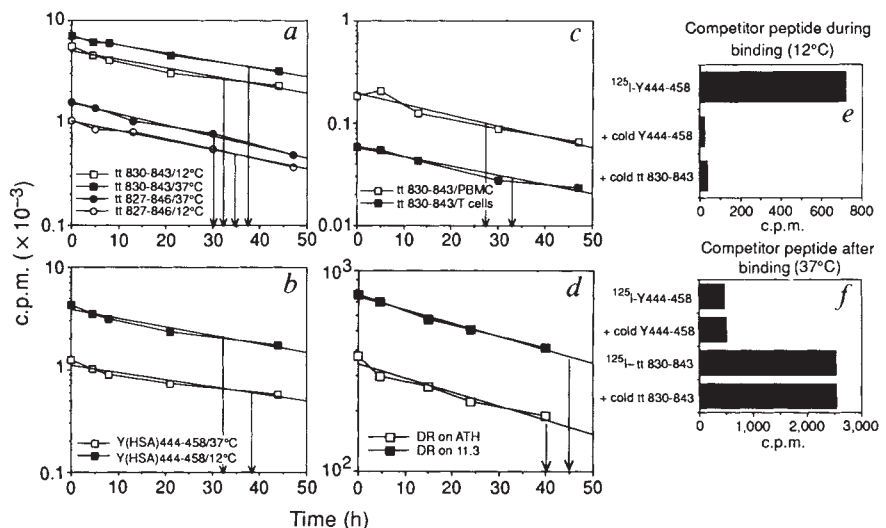
Antigen-specific B cells can stimulate T cells up to 10 days after an initial pulse with antigen¹³. Because the half-life of intact antigen in these B cells is only 2-3 h¹⁴ we investigated whether the persistent stimulatory capacity was due to stable peptide/MHC complexes. We first used T cells to measure the half-life of peptide/MHC complexes in living antigen-presenting cells (APC). Various types of APCs were pulsed with different concentrations of synthetic peptides corresponding to known T-cell epitopes. At different times, the level of stimulatory complexes remaining was measured by titrating the APC in a T-cell proliferation assay¹⁵. A representative experiment (Fig. 1*a-c*) shows that the potency of DRw11⁺ Epstein-Barr virus (EBV)-B cells pulsed with peptide 830-843 from tetanus toxin



(tt) depends on the concentration of peptide and decreases with time of culture. For example, cells pulsed with 1 μ M peptide and then chased for 18, 48 and 96 h gave the same T-cell stimulation when 0.8, 3 and 10×10^3 APC, respectively, were used. To calculate the half-life of the tt830-843/DRw11 complexes we plotted the reciprocal of the number of APC required for a given level of T-cell stimulation against time. This gave a half-life of ~ 25 h and was relatively independent of the level of peptide used for pulsing (Fig. 1*d*). Figure 1*e* shows that the half-life of tt1273-84/DRw52a complexes on EBV-B cells is also ~ 25 h. Similar long half-lives (mean 25 h; range 15-50 h) were recorded in ten different experiments using different APC (EBV-B cells, activated T cells or peripheral blood mononuclear

FIG. 2 Lifetime of peptide/MHC class II complexes in relation to the half-life of class II. *a-c*, Half-life of peptide/class II complexes as determined by pulsing the APC with iodinated peptides and immunoprecipitation of the complexes at different times with anti-DR antibody. *a*, Half-life of complexes between DR molecules and tt830-843 ($\square$, $\blacksquare$) or tt827-846 ($\circ$, $\bullet$) in EBV-B cells that had been pulsed at 37 °C ($\square$, $\bullet$) or 12 °C, pH 5.0 ($\circ$, $\blacksquare$)³⁸. *b*, Half-life of Y444-458-DR complexes on EBV-B cells pulsed at 37 °C ($\square$) or 12 °C, pH 5.0 ($\blacksquare$). *c*, Half-life of tt830-843/DR complexes on PBMC ($\square$) or activated T cells ($\blacksquare$). *d*, Half-life of surface-iodinated DR molecules on EBV-B cells. *e*, Competition for binding to class II DRw11 molecules. *f*, Unlabelled excess peptide does not displace pre-bound labelled peptide.

METHODS. Lifetime of peptide/class II complexes: Peptides were iodinated by the lodogen method⁴⁰ to a specific activity of about 3×10^4 c.p.m. ng⁻¹ and purified on Sephadex G10. EBV-B cells, activated T cells and PBMC (all from DRw11⁺ donors) were pulsed with 2-10 μ M ¹²⁵I-labelled peptide at 37 °C for 4 h ($2-5 \times 10^6$ cells ml⁻¹ in growth medium) or at 12 °C overnight in 140 mM NaCl, 25 mM sodium citrate, pH 5.0, 5 mg ml⁻¹ dialysed BSA. Pulsing was in the presence or absence of a 50-250-fold excess of unlabelled homologous or heterologous peptide. After washing, the cells were divided into aliquots and recultured at a density of $\sim 10^6$ cells ml⁻¹. At different times, class II DR molecules were immunoprecipitated. Associated radioactivity was quantitated by gamma counting. In *a-c*, data have been normalized to give peptide c.p.m. present on 4×10^6 cells. Gel electrophoresis on Tris-tricine SDS-polyacrylamide gels confirmed that the radioactivity recovered from peptide-pulsed cells was in material apparently identical to that used for pulsing. Note that between 30-50% of cell-associated c.p.m. after reculture was class II-associated (that is, immunoprecipitable), demonstrating that large peptide pools do not persist beyond roughly the first hour of reculture. Lifetime of cell-surface DR



molecules: EBV-B cells ($1-2 \times 10^7$) were washed in protein-free medium at 4 °C, resuspended in ~ 50 μ l 0.1 M Na₂HPO₄ and labelled in a final volume of 75 μ l with 1 nmol of either ¹²⁵I-DPSgtc (dithiopropionylsulphosuccinimidylglycyl tyrosylcholamine; refs 22, 23) or ¹²⁵I-labelled sulpho-SHPP (sulphosuccinimidyl-3-(4-hydroxyphenyl) propionate)²¹ prepared as described (ref. 23, and Fig. 3. legend). Cells were washed and recultured in RPMI with FCS for different times. The half-life of surface-labelled DR molecules was calculated either by direct gamma-counting of the washed immunoprecipitates or following excision of the separated α and β chains after SDS-PAGE (Fig. 3). Immunoprecipitations: Cell lysis and immunoprecipitation were done as described^{5,14} using a cocktail of protease inhibitors and the DR-specific monoclonal antibody L243. Reprecipitations from the same lysates showed that the first precipitation was $\sim 70\%$ efficient at all time points.

cells and different peptides that bind to DRw11 (tt830-843, tt947-967, HSA444-458) or to DRw52a (tt1273-84)^{16,17}. Moreover, similar results were obtained when the APC were pulsed with peptide at 0 °C to rule out the possibility that an accumulated intracellular pool of non-MHC-bound peptide was responsible for the long half-life observed (for example, Fig. 1f).

We next measured peptide/class II stability directly by immunoprecipitating class II MHC molecules from cells that had been pulsed with radioiodinated peptides at 37 °C. Labelled peptides (tt830-843 and HSA444-458) remained stably associated with class II molecules on EBV-B cells, activated T cells and PBMC decaying with half-lives between 28 and 38 h (Fig. 2). The somewhat shorter half-lives measured in the functional assay (Fig. 1) may be attributed to cell division. Although tt830-843 and HSA444-456 competed with each other for binding to DRw11 (ref. 17, and Fig. 2e), the stability of preformed complexes did not change when a large excess of either of the unlabelled peptides was added during the chase incubation (Fig. 2f).

To test the possibility that the persistence of the complexes was determined by the lifetime of the class II molecules themselves, we surface-iodinated EBV-B cells and immunoprecipitated class II molecules at different times after reculture at 37 °C. The half-life of the DR molecules was ~40 h (range 30-45 h) on two different EBV lines (Fig. 2d), similar to that already reported for EBV-B cells¹⁸.

The complexes examined so far represent a small fraction of the class II molecules and because these epitopes were selected for their immunogenicity, their behaviour may not reflect the bulk of the class II-bound peptides. Class II $\alpha\beta$ dimers migrating in SDS-polyacrylamide gels contain and are indeed stabilized by bound peptide^{5,19,20}. We reasoned that such peptides might be labelled during cell-surface radioiodination with low-molecular-mass reagents²¹⁻²³. When SDS-stable class II heterodimers, isolated from surface-iodinated cells, were disrupted by electrophoresis into a second gel containing urea, labelled low-molecular-mass material was indeed released from the disrupted $\alpha\beta$ heterodimers (arrow, Fig. 3a) but not (in the second

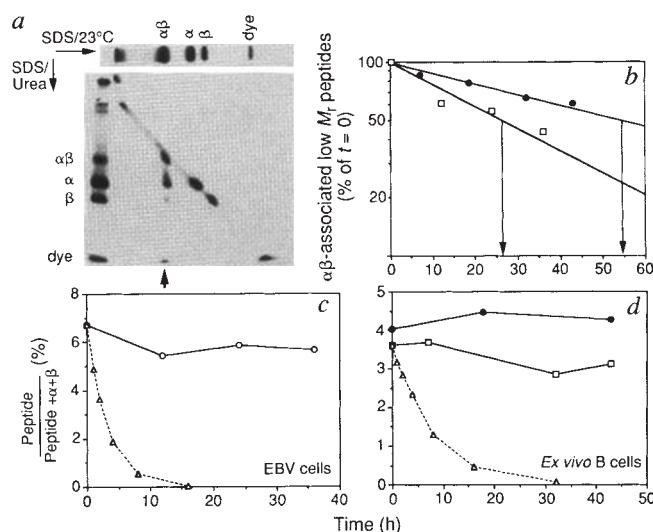
dimension) from the α - and β -chains that separated in the first dimension. This indicated that the material eluted from $\alpha\beta$ dimers could represent endogenous peptides bound to the class II molecules at the time of labelling. We then asked if this material turns over more rapidly than the class II molecules to which it is bound. Labelled class II $\alpha\beta$ complexes were analysed by two-dimensional SDS/SDS+urea gel electrophoresis after different periods of reculture. The results (Fig. 3b) show that the half-life of the low-molecular-mass material was very similar to that of the class II α - and β -subunits themselves, that is, ~25-30 h on EBV-B cells and ~55 h on fresh B cells. Consequently, the ratio of low-molecular-mass labelled material to total label in SDS-stable $\alpha\beta$ dimers remained virtually constant with time (Fig. 3c, d). For comparison we plotted the situation that would be observed if the low-molecular-mass material turned over 10 times faster than the α - and β -chains (dotted lines in Fig. 3c, d). Assuming that we labelled a representative sample of peptides, we conclude that stability is not a special property of selected immunogenic peptides.

Occupancy of class II MHC with peptide induces an SDS-stable conformation^{19,20}. But precisely defined SDS elution conditions may be necessary to preserve all occupied molecules in the dimer state in SDS and these conditions may not be the same for all class II alleles or allele/peptide combinations (refs 19, 20, and R. Germain, personal communication). In fact a significant fraction of the dimers that dissociated in SDS in the first dimension were probably also occupied by peptides (see material of low M_r at end of diagonal in Fig. 3a). The biological half-life of this material and the α - and β -chains running on the diagonal (Fig. 3a) were very similar to that of the SDS-stable α - and β -chains (data not shown), suggesting that in these experiments SDS-stable and SDS-unstable populations may not be functionally distinct.

Our results show that in living human cells class II molecules can bind immunogenic and self peptides irreversibly in the sense that the half-life of the complex is determined by the half-life of the class II molecules themselves. This seems to apply not only to EBV-B cells but also to other APC types such as activated

FIG. 3 Constitutively processed and displayed peptides are stably associated with class II MHC molecules. EBV-B cells (clone A46) or fresh B cells from peripheral blood were surface-iodinated, washed and recultured at 37 °C. At the times indicated, class II MHC molecules were isolated by immunoprecipitation and analysed by two-dimensional SDS/SDS+urea PAGE. a, Typical pattern of an SDS (23 °C)/SDS+urea separation for class II MHC surface-labelled on EBV-B cells with ¹²⁵I-DPSgtc. Note that urea only elutes low- M_r putative peptides (arrowed) from SDS-stable $\alpha\beta$ dimers and not from α and β chains resolved in the first dimension. The less efficient labelling of β chains in SDS-stable dimers suggests a conformational difference relative to SDS-unstable dimers in one or more of the DR alleles precipitated. Shown above the two-dimensional separation is a first dimension SDS (23 °C) separation and to the left a separation only in the second dimension. b, Half-life of low- M_r peptides eluted from SDS-stable class II heterodimers labelled on either EBV cells (□; means of three experiments) or on ex vivo peripheral blood B cells (●; mean of two experiments). c, d, Persistence of low- M_r peptides associated with α and β chains on EBV cells (○) or fresh B cells (●, experiment 1; □, experiment 2) expressed as peptide c.p.m. as a percentage of total $\alpha + \beta$ + peptide c.p.m. for each time point. For comparison, a curve was generated using the equation $\text{pep}_t = \text{pep}_0 \exp [-(k_{\alpha\beta} + k_p)t]$ where pep_t is the fraction of peptide remaining at time t and where class II $\alpha\beta$ decays according to the first-order rate constant $k_{\alpha\beta}$ and peptides 10 times faster, according to k_p . Rate constants were calculated from $k = \ln 2/t_{1/2}$ using a $t_{1/2}$ for class II $\alpha\beta$ of 24 or 60 h on EBV or ex vivo B cells, respectively (b and Fig. 2).

METHODS. B lymphocytes were isolated from fresh mononuclear cells using CD19-coated magnetic beads (Dynal). Cells were resuspended in 0.1 M Na₂HPO₄ at 0 °C and labelled with ¹²⁵I-labelled sulphy-SHPP or DPSgtc as described²¹⁻²³. Briefly, one reaction consisted of 1 nmol SHPP or DPSgtc (1 mM stock in dimethylsulphoxide, -70 °C) added to 14 μ l of a sterile mix of 0.21 M NaCl, 0.14 M sodium phosphate, pH 7.0, 0.36 mg ml⁻¹ chloramine T followed by 10 μ l of ¹²⁵I (1 mCi; Amersham IMS.30). After 20 min, the reaction was stopped by the addition of 1 μ l 1 M sodium *p*-hydroxybenzoate, 0.1 M NaI and the cells were added after 1 min to give a final volume of



~75 μ l. After 20 min the cells were washed in RPMI 1640, 1% FCS and recultured in growth medium. Immunoprecipitations were done with either the DA6.147 or 231 antibodies³⁹ and the proteins eluted from protein A-Sepharose beads by addition of an equal volume of 4% SDS sample buffer at room temperature. Gel slices from the first-dimension 15% gel were excised, equilibrated for 20 min in 0.125 M Tris, pH 6.8, 0.1% SDS, 8 M urea and loaded horizontally onto a 15% gel containing urea. The MHC class II α and β chains and eluted low- M_r peptides were detected by autoradiography, excised and quantified by gamma counting.

T cells, fresh mononuclear cells and B cells (Figs 1 and 3). We cannot exclude the possibility that some class II peptide complexes might dissociate, perhaps during recycling^{23,24}, to generate empty molecules. It will also be important to see whether other class II MHC alleles also bind peptides stably in living cells.

In contrast to isolated peptide/class II MHC complexes, which have been reported to be extremely stable (refs 11, 25–27, and reviewed in ref. 28), estimates for the lifetime of class II MHC/peptide complexes in living cells range from 15 minutes to several hours^{11,12,29–31}. The discrepancy may be a function of the assay used to measure the level of complexes, the diverse peptide/class II molecule combinations and the half-lives of class II MHC in different cell types or species. We suggest however, that persistence of peptide/MHC complexes is likely to be an important aspect of immunogenicity because it will allow antigen encountered over a short period to be presented at a distant site, perhaps several days later. This is well illustrated by dendritic cells^{32,33}, which continue to present epitopes generated *in vivo* or shortly after isolation, but soon lose the capacity to synthesize class II MHC and consequently the ability to present new determinants^{1,2,5}.

Although there is some homology between MHC glycoproteins and the heat-shock protein hsp70 family^{34,35}, the residence time of peptide sequences bound to these two protein families differs dramatically. This is dictated by the ATP-driven, intracellular chaperoning functions of the heat-shock proteins³⁶ on the one hand and on the other by the need for a persistent record of antigen encounter to serve the complex multicellular structure of the immune system. □

Involvement of p21^{ras} in *Xenopus* mesoderm induction

Malcolm Whitman* & D. A. Melton

Department of Biochemistry and Molecular Biology, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

DURING early vertebrate embryogenesis, mesoderm is specified by a signal emanating from prospective endoderm. This signal can respecify *Xenopus* prospective ectoderm as mesoderm, and can be mimicked by members of the fibroblast growth factor and transforming growth factor- β families^{1,2}. In other systems, the p21^{c-ras} proto-oncogene product has been implicated in signal transduction for various polypeptide growth factors. We report here that a dominant inhibitory *ras* mutant blocks the mesoderm-inducing activity of fibroblast growth factor and activin, as well as the endogenous inducing activity of prospective endoderm. A constitutively active *ras* mutant partially mimics these activities. These results indicate that p21^{ras} may have a central role in the transduction of the mesoderm inductive signal. Basic fibroblast growth factor^{3,4} and activin^{5–8} have emerged as candidates for endogenous mesoderm-inducing molecules. The character of the mesoderm induced by these two factors is overlapping but distinct when assessed both by histological and molecular criteria^{9,10}. The signal transduction pathways used during induction by these factors are unknown. We used messenger RNA microinjection of *Xenopus* eggs to express a dominant inhibitory mutant *ras*, p21(Asn 17)^{Ha-ras}, in cells competent to respond to inducing factors to examine the role of p21^{ras} in this response. This mutant, which has a reduced affinity for GTP relative to GDP¹¹, blocks a variety of mitogenic signals in 3T3 fibroblasts¹² as well as the differentiation of pheochromocytoma cells in response to nerve growth factor¹³.

Embryos were injected with RNAs encoding p21^{ras} variants, and at late blastula animal pole tissue (the animal cap) was removed by dissection and treated with activin or basic fibroblast growth factor (bFGF). The earliest observable response of caps to induction by activin and, to a lesser extent, bFGF, is a marked elongation of the cap beginning at the time when undissected sibling embryos gastrulate¹⁴. These movements are drastically reduced in animal caps expressing p21(Asn 17)^{Ha-ras} and treated with activin or bFGF (Fig. 1). Ectopic expression of constitutively active p21^{v-Ha-ras}, without addition of bFGF or activin, causes slight elongation similar to that caused by bFGF (Fig. 1). Animal caps injected with control RNA encoding a transformation-defective point mutant of p21^{v-Ha-ras}, p21(Ser 34)^{v-Ha-ras}, do not elongate (Fig. 1), nor do caps from uninjected embryos (not shown). After 48 h, animal caps were examined for mesodermal structures (Table 1). Control injected caps treated with activin contain blocks of muscle, notochord and neural structures including eyes, neural tubes, and brain vesicles¹⁵. Caps from embryos expressing p21(Asn 17)^{Ha-ras}, by contrast, consist only of atypical epidermis characteristic of uninduced caps. Caps expressing constitutively active p21^{v-Ha-ras} show some induction, indicated by the appearance of mesothelium and small numbers of muscle cells, but do not form discrete mesodermal structures (not shown). As little as 200 pg per embryo of p21(Asn 17)^{Ha-ras} RNA was found to be sufficient to block activin induction, and as little as 20 pg of p21^{v-Ha-ras} RNA was sufficient to induce mesodermal markers. Coinjection of a 20-fold excess of p21(Asn 17)^{Ha-ras} RNA over p21^{v-Ha-ras} RNA did not prevent induction by p21^{v-Ha-ras} (not shown), indicating that p21^{v-Ha-ras} acts on targets downstream to the p21(Asn 17)^{Ha-ras}-generated block to induction.

Received 12 December 1991; accepted 3 April 1992.

1. Adorini, L., Ullrich, S. J., Appella, E. & Fuchs, S. *Nature* **346**, 63–67 (1990).
2. St Pierre, Y. & Watts, T. H. *J. Immunol.* **145**, 812–818 (1990).
3. Neefjes, J. J., Stollorz, V., Peters, P. J., Geuze, H. J. & Ploegh, H. L. *Cell* **61**, 171–183 (1990).
4. Peters, P. J., Neefjes, J. J., Oorschot, V., Ploegh, H. L. & Geuze, H. J. *Nature* **349**, 669–676 (1991).
5. Davidson, H. W., Reid, P. A., Lanzavecchia, A. & Watts, C. *Cell* **67**, 105–116 (1991).
6. Teyton, L. *et al.* *Nature* **348**, 39–44 (1990).
7. Germain, R. N. & Hendrix, L. R. *Nature* **353**, 134–139 (1991).
8. Roche, P. A. & Cresswell, P. *Nature* **345**, 615–618 (1990).
9. Neefjes, J. J. & Ploegh, H. L. *EMBO J.* **11**, 411–416 (1992).
10. Pernis, B. *Immun. Today* **6**, 45–49 (1985).
11. Harding, C. W., Roof, R. W. & Ulanue, E. R. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4230–4234 (1989).
12. Adorini, L., Appella, E., Doria, G., Cardinaux, F. & Nagy, Z. A. *Nature* **342**, 800–803 (1990).
13. Lanzavecchia, A. *Immunol. Rev.* **99**, 39–51 (1987).
14. Davidson, H. W. & Watts, C. *J. Cell Biol.* **109**, 85–92 (1989).
15. Mathis, L. A., Glimcher, L. H., Paul, W. E. & Schwartz, R. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6019–23 (1983).
16. Panina-Bordignon, P. *et al.* *Eur. J. Immunol.* **19**, 2237–2242 (1989).
17. Panina-Bordignon, P., Corradin, G., Roosnek, E., Sette, A. & Lanzavecchia, A. *Science* **252**, 1548–1550 (1991).
18. Davis, J. & Cresswell, P. *J. Immunol.* **144**, 990–997 (1990).
19. Sadegh-Nasseri, S. & Germain, R. N. *Nature* **353**, 167–170 (1991).
20. Stern, L. J. & Wiley, D. C. *Cell* **68**, 465–477 (1992).
21. Thompson, J. A., Lav, A. L. & Cunningham, D. D. *Biochemistry* **26**, 743–750 (1983).
22. Bretscher, M. S. & Lutter, R. *EMBO J.* **7**, 4087–4092 (1988).
23. Reid, P. A. & Watts, C. *Nature* **346**, 655–657 (1990).
24. Harding, C. V. & Ulanue, E. R. *J. Immunol.* **142**, 12–19 (1989).
25. Jensen, P. E. *J. Immunol.* **143**, 420–425 (1989).
26. Lee, J. M. & Watts, T. H. *J. Immunol.* **144**, 1829–1834 (1990).
27. Roche, P. & Cresswell, P. *J. Immunol.* **144**, 1849–56 (1990).
28. Rothbard, J. B. & Geffer, M. L. A. *Rev. Immunol.* **9**, 527–565 (1991).
29. Nygard, N. R. *et al.* *J. exp. Med.* **174**, 243–251 (1991).
30. Busch, R., Strang, G., Howland, K. & Rothbard, J. B. *Int. Immunol.* **2**, 443–451 (1990).
31. Pedrazzini, T., Sette, A., Albertson, M. & Grey, H. M. *J. Immunol.* **146**, 3496–3501 (1991).
32. Puré, E. *et al.* *J. Exp. Med.* **172**, 1459–1469 (1990).
33. Kampgen, E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 3014–3018 (1991).
34. Rippmann, F., Taylor, W. R., Rothbard, J. B. & Green, N. M. *EMBO J.* **10**, 1053–1059 (1991).
35. Flajnik, M. F., Canel, C., Kramer, J. & Kasahara, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 537–541 (1991).
36. Flynn, G. C., Pohl, J., Flocco, M. T. & Rothman, J. E. *Nature* **353**, 726–730 (1991).
37. Lanzavecchia, A. *Nature* **314**, 537–539 (1985).
38. Jensen, P. *J. exp. Med.* **171**, 1779–1784 (1990).
39. Guy, K., Van Heyningen, V., Cohen, B. B., Dean, D. L. & Steel, C. M. *Eur. J. Immunol.* **12**, 942–948 (1982).
40. Fraker, P. J. & Speck, J. C. *Biochem. biophys. Res. Commun.* **80**, 849–857 (1978).

ACKNOWLEDGEMENTS. We thank G. P. Corradin and A. Sette for peptides, S. Moore, S. Demotz, K. Karjalainen and S. Abgrinani for critical reading of the manuscript and R. N. Germain for comments. This work was supported in part by the Medical Research Council and the Wellcome Trust. The Basel Institute for Immunology was founded and is supported by F. Hoffmann La Roche & Co. Ltd, Basel, Switzerland.

* Present address: Department of Cellular and Molecular Physiology, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA.

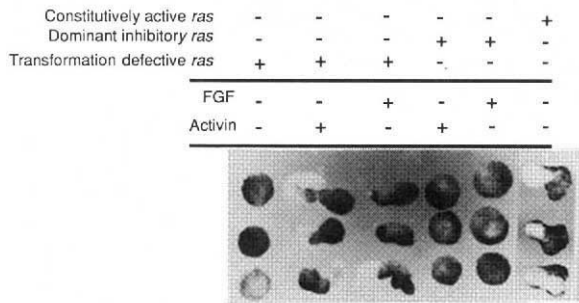


FIG. 1 Effects on elongation of animal caps expressing dominant inhibitory *ras* after induction. Animal caps were cut from embryos injected with RNA encoding non-transforming p21(Ser 34)^{Ha-ras}, dominant inhibitory p21(Asn 17)^{Ha-ras}, or constitutively active p21^{v-Ha-ras} at stage 8.5 and were either untreated or treated with activin or FGF. Three caps are shown for each condition. The elongation movements that accompany induction by peptide growth factors mimic the convergent extension of mesoderm during normal development¹⁴.

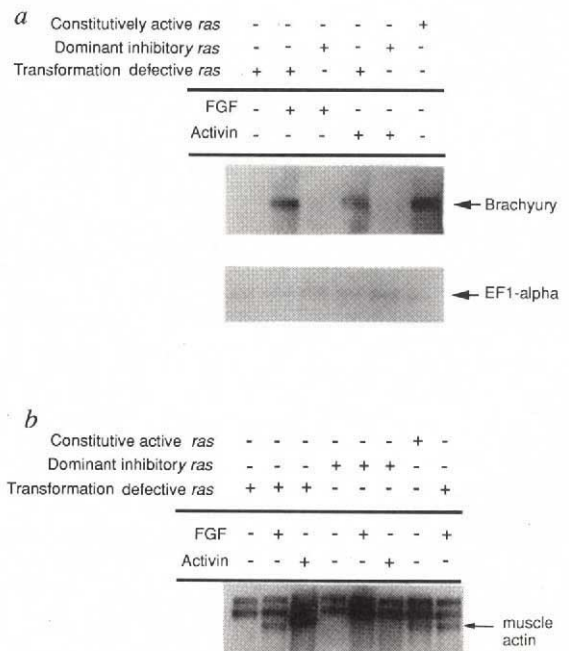
METHODS. Eggs were obtained from *Xenopus laevis* females and embryos were raised as described²⁷. Both blastomeres of two-cell embryos were microinjected with 0.2–2 ng mRNA prepared by transcription *in vitro*²⁸. RNAs encoding p21(Asn 17)^{Ha-ras} and p21(Ser 34)^{Ha-ras} were transcribed from complementary DNAs in the vector pSP64T, which contains globin untranslated sequences that enhance protein expression from injected RNA²⁷; RNA encoding p21^{v-Ha-ras} was transcribed in the pGEM 7 expression vector. At the doses used, none of these RNAs inhibited early or late cleavages, although at higher doses RNA encoding p21^{v-Ha-ras} markedly inhibited cleavage as previously described²⁰. At stage 8.5, animal caps were dissected and treated with 3 nM FGF or 100 pM activin. After 8–10 h in salt solution (0.5 × MMR)²⁶ caps were examined for elongation by microscopy. Results shown are typical for *n* = 50 caps per condition.

To characterize further the effect of p21^{ras} mutants on induction, early and late molecular markers for mesoderm were examined. *Xenopus* brachyury, an immediate early marker for mesoderm¹⁶, is expressed in the marginal zone of intact gastrulae and is rapidly induced in animal caps by FGF and activin. Expression of the dominant inhibitory p21(Asn 17)^{Ha-ras} in animal caps completely prevents induction of brachyury by FGF or activin (Fig. 2a). Expression of constitutively active p21^{v-Ha-ras} alone is sufficient to induce brachyury in animal caps to levels

similar to those with FGF or activin. Figure 2b shows expression of muscle-specific actin, a late marker for mesodermal differentiation. Animal caps injected with control RNA express cytoskeletal actin mRNAs, but no mesoderm-specific muscle actin. When caps are treated with FGF or activin, muscle actin mRNA expression is induced as previously shown^{4,5}. Expression of muscle actin is completely blocked in FGF-treated, activin-treated or untreated control caps taken from embryos injected with RNA encoding p21(Asn 17)^{Ha-ras}. Caps from embryos injected with mRNA encoding constitutively active p21^{v-Ha-ras} expressed muscle actin in the absence of FGF or activin treatment. Expression of a marker characteristic of ectodermal differentiation, the epidermal cytokeratin XK70 (ref. 17), was unaffected in caps from p21(Asn 17)^{Ha-ras}-injected embryos relative to caps from control injected embryos (not shown). Expression of p21(Asn 17)^{Ha-ras} did not block the ability¹⁸ of microinjected *xwnt-8* RNA to induce a second dorsal blastopore lip (C. LaBonne and M. W., unpublished results), nor did it interfere with prostaglandin E₂ stimulation of adenylate cyclase in animal caps from early embryos (data not shown). This indicates that p21(Asn 17)^{Ha-ras} does not produce a generalized block to receptor-activated signal transduction but rather selectively blocks events mediating mesoderm inductive signals.

Because of the dramatic effect of p21(Asn 17)^{Ha-ras} on induction by peptide factors, we examined whether p21(Asn 17)^{Ha-ras} would block the endogenous inducing activity of prospective endoderm. Animal caps expressing defective or dominant inhibitory p21^{ras} were combined with vegetal bottoms from uninjected embryos (Nieuwkoop recombinants)¹⁹. Recombinants made with caps from control injected embryos showed extensive elongation at gastrulation (Fig. 3a), expression of muscle actin (Fig. 3b), and differentiation of mesodermal structures (Table 1). Recombinants made with caps expressing p21(Asn 17)^{Ha-ras} showed little or no indications of mesodermal induction by any of these criteria (Fig. 3a, b; Table 1). Recombinants made with animal caps from uninjected embryos and vegetal bottoms from embryos expressing p21(Asn 17)^{Ha-ras} showed indications of mesoderm induction indistinguishable from those made using vegetal bottoms from control injected embryos. These results indicate that p21(Asn 17)^{Ha-ras} blocks the capacity of animal caps to respond to the endogenous inducing signal of the vegetal pole.

FIG. 2 The effects of *ras* on induction of mesoderm markers. Animal caps from embryos injected after fertilization with mRNA encoding non-transforming p21(Ser 34)^{ras}, dominant inhibitory p21(Asn 17)^{ras}, or constitutively active p21^{v-Ha-ras} were cut at stage 8.5 and untreated or treated with 3 nM FGF or 100 pM activin. mRNA was extracted 3 h (a) or 36 h (b) later for Northern blot analysis. a, Northern blot analysis using an antisense RNA probe for *Xenopus* brachyury. As a control for RNA loading, blot was stripped and reprobed for EF1-alpha (bottom). b, Northern blot analysis using an antisense RNA probe for *Xenopus* cardiac actin. The doublet above cardiac actin is crossreacting cytoskeletal actin²⁹.



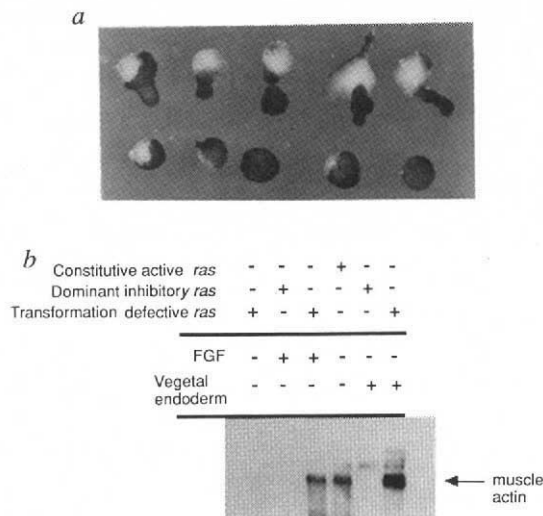


FIG. 3 Effects of *ras* on Nieuwkoop recombinants. Animal caps from embryos injected with RNA encoding *ras* mutants were combined with vegetal bottoms and examined for elongation (a) of muscle actin expression (b) as described above. a, Animal caps expressing non-transforming p21(Ser 34)^{ras} (top row) or dominant inhibitory p21(Asn 17)^{ras} (bottom row) were recombined with uninjected vegetal bottoms and examined for elongation. Results shown are typical for *n*=20 caps per condition. b, Animal caps expressing non-transforming p21(Ser 34)^{ras}, dominant inhibitory p21(Asn 17)^{Ha-ras}, or constitutively active p21^{v-Ha-ras} were cut at stage 8.5 and not treated, treated with 3 nM FGF, or recombined with vegetal bottoms, cultured, processed and probed with a cardiac actin probe as described in Fig. 2b legend.

Undissected embryos injected with p21(Asn 17)^{Ha-ras} mRNA at first cleavage appear to develop normally through mid-gastrula, but fail to complete gastrulation and later develop severe posterior axial defects. Because of the disruption in gastrulation, it is difficult to interpret the subsequent defects in terms of effects of p21(Asn 17)^{Ha-ras} on pre-gastrulation specification of mesodermal pattern. Ectopic expression of p21^{v-Ha-ras} at levels sufficient to induce mesoderm in animal caps disrupted gastrulation (not shown); higher levels of expression inhibited early cleavage²⁰.

Induction by activin, FGF or endoderm seems to require a signalling step which can be blocked by a dominant inhibitory mutant of p21^{ras}. This provides direct evidence that the endogenous vegetally localized mesoderm-inducing activity identified by Nieuwkoop¹⁹ shares a membrane signal transduction pathway with the recently characterized polypeptide mesoderm-inducing factors. It is not known to what extent different members of the *ras* family share overlapping signalling pathways²¹, and it is therefore premature to conclude that p21^{v-Ha-ras}, rather than another member of the *ras* family, is of central importance in inductive signalling. Overexpression of transforming p21^{v-Ha-ras} induces some mesodermal molecular markers and tissues, indicating that activation of p21^{ras} is sufficient to mimic inductive signals. It is not, however, sufficient

for the specification of all mesodermal tissues which may require more information than the activation of a single signal transduction pathway.

Although p21^{ras} has been postulated to function as a coupling protein between transmembrane receptors and cellular effectors^{22,23}, little evidence is available concerning its regulation or targets. Previous studies using p21(Asn 17)^{Ha-ras} have indicated that p21^{ras} participates in the transduction of the mitogenic effects of tyrosine kinase growth factor receptors^{12,13}. We have previously found that activation of a family of cellular tyrosine kinases by expression of the viral oncogene polyoma middle T is sufficient to induce mesoderm²⁴. The inhibitory action of p21(Asn 17)^{Ha-ras} on the mesoderm-inducing activity of FGF, which activates a receptor tyrosine kinase²⁵, is consistent with this earlier work. More surprising, however, is the inhibitory effect of p21(Asn 17)^{Ha-ras} on the action of activin. Little is known about the early steps in the transduction of the activin signal, but the recent cloning of an activin receptor indicates that it is a Ser-Thr specific transmembrane protein kinase²⁶. Our observations indicate that in early embryos activin may share a *ras*-dependent signalling step with FGF. □

TABLE 1 Effects of dominant inhibitory *ras* on embryonic induction

	Notochord	Muscle	Neural*
Control injection	0/11	0/11	0/11
Control injection +activin	9/11	10/11	4/11
Dominant inhibitory injection +activin	0/7	0/7	0/7
Control injection +vegetal bottom	2/9	9/9	0/9
Dominant inhibitory injection +vegetal bottom	0/11	0/11	0/9

Dominant inhibitory p21^{ras} blocks induction of mesodermal and neural structures by activin and endoderm. Animal caps injected with control (transformation defective p21(Ser 34)^{Ha-v-ras} RNA) or dominant inhibitory p21(Asn 17)^{Ha-ras} RNA were dissected and either treated with activin or recombined with vegetal bottoms from uninjected embryos. After 72 h, caps or recombinants were fixed in 4% formaldehyde, 0.1 M cacodylate pH 7, embedded in paraplast, sectioned at 7 µm and stained with Giemsa. Sections were scored by histological inspection for blocks of muscle tissue, notochord, or neural structures.

* Positive scoring for neural structures required the presence of neural tube, brain vesicle, or eyes, not merely the presence of darkly stained, compact neural tissue.

Received 7 February; accepted 2 April 1992.

- Smith, J. C. *Development* **105**, 665-667 (1989).
- Whitman, M. Melton, D. A. *Rev. Cell Biol.* Vol. 5, 93-117 (1989).
- Slack, J. M. W., Darlington, B. G., Heath, J. K. & Godsave, S. F. *Nature* **326**, 197-200 (1987).
- Kimelman, D., Abraham, J. A., Haaparanta, T., Palishi, T. M. & Kirschner, M. W. *Science* **243**, 1053-1057 (1988).
- Smith, J. C., Price, B. M. J., Van Nimmen, K. & Huylebroeck, D. *Nature* **345**, 729-731 (1990).
- Thomsen, G. *et al. Cell* **63**, 485-493 (1990).
- Asashima, M. *et al. Wilhelm Roux Arch. dev. Biol.* **198**, 330-335 (1990).
- Mitrani, E. *et al. Cell* **63**, 495-501 (1990).
- Ruiz i Altaba, A. & Melton, D. A. *Nature* **341**, 33-38 (1989).
- Godsave, S. F., Isaacs, H. V. & Slack, J. M. W. *Development* **102**, 555-566 (1988).
- Feig, L. A. & Cooper, G. M. *Molec. cell. Biol.* **8**, 3235-3243 (1988).
- Cai, H., Szeberenyi, J. & Cooper, G. M. *Molec. cell. Biol.* **10**, 5314-5323 (1990).
- Szeberenyi, J., Cai, H. & Cooper, G. M. *Molec. cell. Biol.* **10**, 5324-5332 (1990).
- Symes, K. & Smith, J. *Development* **101**, 339-345 (1987).
- Sokol, S., Wong, G. G. & Melton, D. A. *Science* **249**, 561-564 (1990).
- Smith, J. C. *Cell* **67**, 79-88 (1991).
- Jamrich, M., Sargent, T. & Dawid, I. B. *Genes Dev.* **1**, 124-132 (1987).
- Sokol, S., Christian, J. L., Moon, R. T. & Melton, D. A. *Cell* **67**, 741-752 (1991).
- Nieuwkoop, P. D. *Wilhelm Roux Arch. EntwMech. Org.* **162**, 341-373 (1969).
- Daar, I. *et al. Science* **253**, 74-76 (1991).
- Hall, A. *Science* **249**, 635-640 (1990).
- Barbacid, M. A. *Rev. Biochem. Sci.* **56**, 779-827 (1987).
- Hall, A. *Cell* **61**, 921-923 (1990).
- Whitman, M. & Melton, D. A. *Science* **244**, 803-806.
- Lee, P. L., Johnson, D. E., Cousens, L. S., Fried, V. A. & Williams, L. T. *Science* **245**, 57-60 (1989).
- Mathews, L. S. & Vale, W. W. *Cell* **65**, 973-982 (1991).
- Newport, J. & Kirschner, M. *Cell* **30**, 675-686 (1982).
- Krieg, P. A. & Melton, D. A. *Meth. Enzym.* **155**, 397-415 (1987).
- Gurdon, J. B., Fairman, S., Mohun, T. J. & Brennan, S. *Cell* **41**, 913-922 (1985).

ACKNOWLEDGEMENTS. We thank C. Hong and G. Cooper for p21^{ras} cDNAs, P. Wilson for *Xenopus* brachyury probe, T. Sargent for anti-XK70 antibody, C. LaBonne for examining the effect of p21(Asn 17)^{Ha-ras} on *xwnt-8*, and P. Klein for critical reading of the manuscript. M.W. is a scholar of the Lucille P. Markey Charitable Trust. This work was also supported by a grant from NIH.

Induction of gut in *Caenorhabditis elegans* embryos

Bob Goldstein

Center for Developmental Biology, Department of Zoology,
University of Texas, Austin, Texas 78712, USA

TWO types of developmental events can cause an embryonic cell to adopt a fate different from that of its neighbours: during a cell division particular contents may be segregated to only one daughter cell and cells may experience different external cues, commonly in the form of inductive cell interactions. Work on development in the nematode *Caenorhabditis elegans* suggests that most cell fates are specified without a need for cell interactions. In particular, the gut cell lineage of *C. elegans* has been used as a primary example of specification by differential segregation of determinants¹. Here I re-examine the role of induction in gut specification by isolating early blastomeres. In *C. elegans*, the gut derives from all the progeny of a single blastomere (E) of the eight-cell stage². When a gut precursor cell (EMS) is isolated during the first half of the four-cell stage, gut does not differentiate. Gut differentiation is rescued by recombining EMS with its posterior neighbour (P₂), but not by recombining EMS with one or both of the other two cells of the four-cell stage. These results demonstrate that P₂ induces EMS to form gut in *C. elegans*.

Cells were isolated from the influence of their neighbours by removing egg shells from cleavage-stage embryos and physically separating cells from each other (Fig. 2 legend). After isolating cells during the four-cell stage, gut often differentiated from the EMS isolate (21 out of 40 cases) but never from the P₂, ABa or ABp isolates (Figs 1 and 2). An investigation of the timing during the four-cell stage of each isolation suggests that EMS may need to contact one or more of its neighbours early in the four-cell stage for gut to differentiate: EMS cells isolated during the first half of the four-cell stage never differentiated gut, whereas those isolated during the second half of the four-cell stage often did (Fig. 3a). In EMS isolates that did not differentiate gut, the normal slowing of E-cell cycles after one division did not occur; the E lineage divided synchronously with the MS lineage (Fig. 4). This suggests that in the absence of the induction E may take on the fate of MS; however, differentiation of other tissues has not yet been assayed in uninduced EMS cells. This alteration in the timing of the E-lineage cell cycles also occurs in mutants that do not form gut (J. Shaw and D. Morton, personal communication.)

The finding that gut will not differentiate if EMS is isolated early in the four-cell stage suggests that during the four-cell stage a cell interaction occurs that specifies gut in EMS. Alternatively, separating cells from each other early in the cell cycle may damage cells so that although EMS continues to cleave, it will not form gut. To distinguish between these two possibilities,

cells were separated from each other during the first half of the four-cell stage, and then within one to two minutes all four cells were placed back in contact with each other. If separating cells simply damages EMS, then gut should not differentiate. If, however, separating cells prevents them from interacting to specify gut, then recombining cells should allow the interaction to continue, and gut should differentiate. In such an experiment, gut differentiates (5 out of 5 cases). Thus EMS requires contact with one or more of the other cells during the four-cell stage in order for its E lineage to differentiate gut.

To determine which of the other three cells induce gut in EMS, cells were separated during the four-cell stage and EMS was then recombined with either P₂, or one or both of ABa and ABp. When cells were separated from each other during the four-cell stage and EMS was immediately placed back in contact with one or both of ABa and ABp, gut differentiation was not rescued (Fig. 3b). When cells were separated from each other during the four-cell stage and EMS immediately placed back in contact with P₂, gut differentiated in every case (Fig. 3c). These experiments demonstrate that P₂, which normally sits at the posterior end (the 'E' end) of EMS, induces EMS to form gut in its E lineage.

These results should not be misconstrued as proof that segregation of determinants plays no part in gut specification. Although it remains to be tested, it is possible that segregation may localize gut potential to EMS, and the induction may then act as a positional cue to make the two daughter cells of EMS differentiate along different pathways. These experiments do, however, demonstrate that there is an induction during the four-cell stage that is necessary for specification of gut.

It has previously been suggested³ that an interaction between P₂ and EMS specifies gut: out of six cases in which a P₂ cell was removed from a four-cell embryo, in none did gut differentiate. It was not clear, however, whether removing P₂ prevented an interaction with EMS or simply damaged EMS, and in a similar experiment⁴ gut in fact usually differentiated.

The original suggestion that gut specification occurs independently of induction derives from experiments in which cells were isolated by applying pressure on embryos and then individuals were picked out in which all cells except the cell of interest were lysed^{5,6}. When the E cell or its precursors were isolated by this method at the two-, four-, or eight-cell stage, gut often differentiated^{5,6}. This result suggested that gut is specified independently of cell interactions and thus solely by virtue of determinants segregated into E during the divisions leading to its formation. But because the time in the cell cycle when cells were isolated was not noted, this method could not detect the type of interaction described here, which occurs between sister cells soon after their formation.

Several cells are specified by cell interactions during late embryonic and post-embryonic development in *C. elegans* (reviewed in ref. 7). Interactions are also necessary during early embryonic development to specify the fates of several cells in

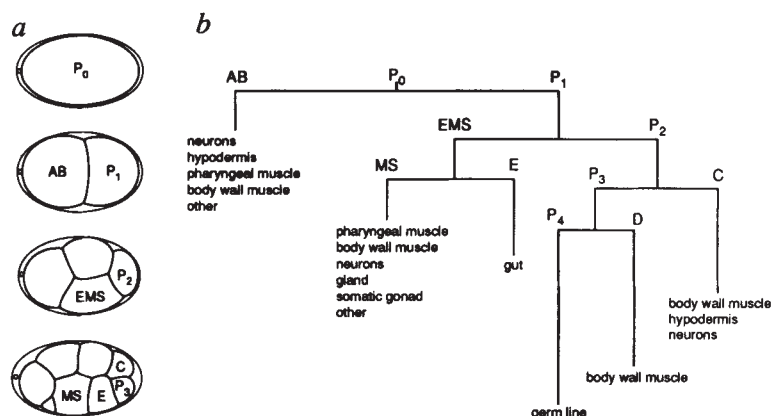
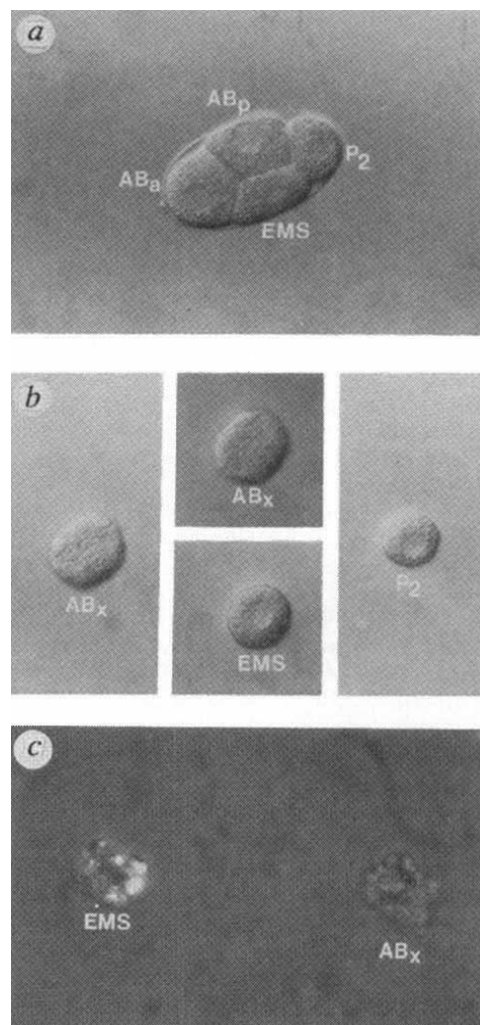


FIG. 2 Cell isolation at the four-cell stage. *a*, Intact four-cell stage embryo. *b*, Four cells after isolation. ABa and ABp are labelled ABx after the separation as these cannot be distinguished from each other. Nuclei have broken down in ABx cells in preparation for cell division. *c*, EMS and ABx isolates 12 h later under polarized light. Birefringent rhabditi granules have developed in the EMS isolate but not in the ABx isolate.

METHODS. Animals were maintained by standard techniques¹¹ at 21.5 °C. Gravid hermaphrodites were cut in egg salts⁶ on a depression slide to release embryos. Early cleavage stages were collected and transferred to a 15% solution of hypochlorite (Aldrich) in egg salts. After 3 min, embryos were transferred through two changes of embryonic growth medium (EGM, developed by L. Edgar, personal communication; methods available on request) and then to a solution of 20 units per ml chitinase (Sigma), 20 mg ml⁻¹ *a*-chymotrypsin (Sigma, type II), 1% penicillin-streptomycin solution (Gibco). After 6–9 min, as the eggshells softened and cells began to round up, an equal volume of trypsin inhibitor (10 mg ml⁻¹) (Worthington) in EGM was added. After 30 s, embryos were transferred through two more changes of EGM. The vitelline membrane was removed in EGM by drawing each embryo in and out of a pulled glass needle (glass capillary tubes, A-M Systems) whose tip had been cut to an open diameter slightly shorter than the short axis of the embryo. Capillary tubes were siliconized in Prosil-28 (PCR, Inc.) before pulling. Embryos allowed to develop unperturbed from this point differentiate gut in 92% of all cases ($n=57/62$). Cells were separated from each other as a result of the turbulence created by drawing each embryo several times in and out of this needle or one of a slightly larger tip diameter. In about half of these cases, two or more cells lysed and the embryo was discarded; the remainder were used. Cells removed from the egg shell invariably divided in the normal order. In more than 95% of all cases, ABa and ABp divided synchronously, EMS divided 3–4 min later, and P₂ divided 3–5 min after EMS. Thus after separating cells and identifying them on the basis of size, division order was a useful check to see that no mistakes had been made. Embryos were disaggregated one at a time and were kept separately to avoid confusing cells. The time in the cell cycle when EMS was isolated was established by recording the time of separation and then the time at which cytokinesis began in EMS. As only some embryos in each batch can be used because many are damaged, and manipulations must be rapid, it is not feasible to record the time of P₁ cleavage as well for every embryo. This time has been at least estimated for every embryo: while preparing to remove egg shells, embryos were arranged in order of age as AB and P₁ divided in each embryo. This method allows one to know when P₁ divides in a few embryos, and then to recognize embryos that develop much more slowly or quickly than the few recorded. In this way the EMS cell cycle in disaggregated embryos can be estimated at about 15 to 20 min, compared with about 15 min for embryos inside their eggshells. In the cleavages that follow, cells continue to divide in the proper order, only up to twice as slowly as embryos inside their eggshells. Differentiation of



gut was scored by checking for the presence of rhabditi granules, a gut-specific differentiation marker^{5,12}, under polarized light at 8–24 h after first cleavage.

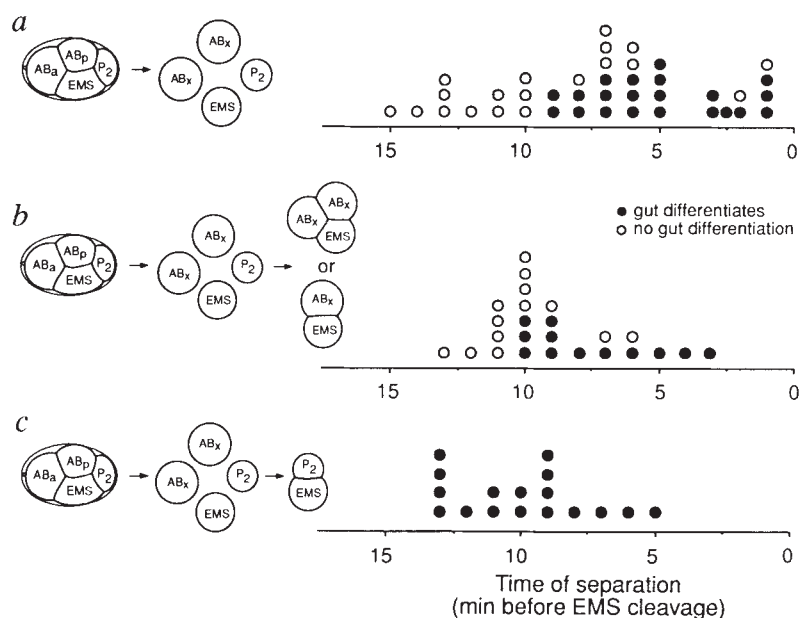


FIG. 3 Results of isolation and recombination experiments. Figures on the left show the experiments and histograms on the right show the results. Each circle in the histograms represents one case; cases are stacked vertically over the time at which each separation was done. Time lines each represent the entire four-cell stage. Time of separation was determined as described in Fig. 2 legend. *a*, Gut differentiation after isolating EMS at various times throughout the four-cell stage. *b* and *c*, Gut differentiation after isolating EMS at various times throughout the four-cell stage and then (within 1–2 min) recombining EMS with either one or both of ABa and ABp (*b*) or with P₂ (*c*). ABa and ABp are both labelled ABx after a separation, as these cannot be distinguished from each other. Cases where EMS was recombined with only one ABx cell and gut did not differentiate are at 12 min (one case) and 11 min (two cases) before EMS cleaved, and at 10 (three cases), 9 (one case), 8 (one case), 5 (one case) and 3 (one case) min before EMS cleaved in which gut did differentiate. In all other cases in *b*, EMS was recombined with both ABx cells. As expected, in both recombination experiments (*b*, *c*) gut differentiated in most cases if the separation was done late in the four-cell stage, as results in *a* indicate that gut is induced by this time. Thus the results from separation and recombination experiments during the first half of the four-cell stage indicate whether or not, after isolating EMS, gut differentiation was rescued by the recombination.

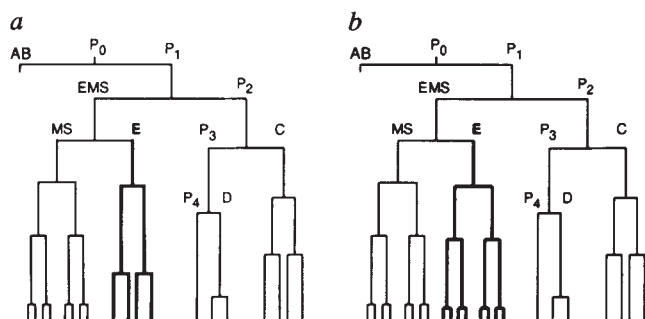


FIG. 4 *a*, Partial lineage of intact embryo. *b*, Partial lineage of embryos in which P₂ and EMS were separated and gut did not differentiate, showing the relative order of division of cells. The E lineage takes on MS-like cleavage times when the induction is blocked. (9/9 cases observed, from the 19 that did not differentiate gut, shown in Fig. 3a.)

the AB lineage^{4,8,9}. No other lineages were previously known to require cell interactions during early development.

Two pieces of information raise the possibility that many other inductions may be occurring during early *C. elegans* development. First, this work shows that previous methods could not detect all cell interactions. Second, most issues in *C. elegans* arise from many distantly related cells, not from the progeny of a single founder cell² (Fig. 1*b*); segregation of determinants

would have to be extraordinarily complex to generate this sort of lineage independently of inductions, considering that determinants for at least several cell types are not prelocalized in the egg¹⁰.

Given the organism's amenability to genetic analysis, the finding that gut is induced and the possibility that many other inductions may be occurring in *C. elegans* may afford us a chance to understand the molecular mechanisms that underlie embryonic inductions. □

Received 7 November 1991; accepted 7 April 1992.

1. Davidson, E. H. *Gene Activity in Early Development* 3rd edn (Academic, Orlando, 1986).
2. Sulston, J. E. *et al. Devl Biol.* **100**, 64–119 (1983).
3. Schierenberg, E. *Devl Biol.* **122**, 452–463 (1987).
4. Priess, J. R. & Thomson, J. N. *Cell* **48**, 241–250 (1987).
5. Lauffer, J. S. *et al. Cell* **19**, 569–577 (1980).
6. Edgar, L. G. & McGhee, J. D. *Devl Biol.* **114**, 109–118 (1986).
7. Maine, E. M. & Kimble, J. *BioEssays* **12**, 265–271 (1990).
8. Wood, W. B. *Nature* **349**, 536–538 (1991).
9. Schnabel, R. *Mech. Dev.* **34**, 85–100 (1991).
10. Lauffer, J. S. & von Ehrenstein, G. *Science* **211**, 402–405 (1981).
11. Sulston, J. & Hodgkin, J. in *The Nematode Caenorhabditis elegans* (ed. Wood, W. B.) 587–606 (Cold Spring Harbor Laboratory Press, New York, 1988).
12. Chitwood, B. G. & Chitwood, M. B. *Introduction to Nematology* (University Park, Baltimore, 1974).

ACKNOWLEDGEMENTS. I thank G. Freeman for advice, C. Jenkins, G. Freeman and others for critically reading the manuscript, and L. Edgar for protocols for isolating and culturing blastomeres. Worms were provided by D. Barker and by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources. This work was supported by an NSF grant to Gary Freeman and an NIH predoctoral training grant to the author.

The rate of actin-based motility of intracellular *Listeria monocytogenes* equals the rate of actin polymerization

Julie A. Theriot*, Timothy J. Mitchison†, Lewis G. Tilney‡ & Daniel A. Portnoy§

Departments of *Biochemistry and Biophysics and †Pharmacology, University of California, San Francisco, California 94143, USA

‡Department of Biology, University of Pennsylvania and

§Department of Microbiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA

THE Gram-positive bacterium *Listeria monocytogenes* is a facultative intracellular pathogen capable of rapid movement through the host cell cytoplasm¹. The biophysical basis of the motility of *L. monocytogenes* is an interesting question in its own right, the answer to which may shed light on the general processes of actin-based motility in cells. Moving intracellular bacteria display phase-dense 'comet tails' made of actin filaments, the formation of which is required for bacterial motility^{2,3}. We have investigated the dynamics of the actin filaments in the comet tails using the technique of photoactivation of fluorescence, which allows monitoring of the movement and turnover of labelled actin filaments after activation by illumination with ultraviolet light. We find that the actin filaments remain stationary in the cytoplasm as the bacterium moves forward, and that length of the comet tails is linearly proportional to the rate of movement. Our results imply that the motile mechanism involves continuous polymerization and release of actin filaments at the bacterial surface and that the rate of filament generation is related to the rate of movement. We suggest that actin polymerization provides the driving force for bacterial propulsion.

L. monocytogenes can move in many cell types, and we chose to infect the flat and easily injectable pectoral kidney epithelial (PtK2) cell line. As reported¹, *L. monocytogenes* induced the formation of actin-filament-rich phase-dense comet tails and

moved at rates of up to 0.4 μm per second through the cytoplasm of these cells. The tails appear substantially shorter by phase contrast than by rhodamine-phalloidin labelling, which marks filamentous actin (Fig. 1*a, b*). Fluorescence intensity profiles of phalloidin-labelled tails reveal that there is a pronounced gradient of actin filament density through the tail, such that the filament density is highest closest to the bacterium and decreases exponentially towards the distal top (Fig. 1*c*). The tail is visible by phase microscopy as long as 30–50% of the peak density of actin filaments in the tail persist.

To probe *L. monocytogenes* tail actin filament dynamics, infected cells were microinjected with purified rabbit skeletal muscle G-actin covalently coupled to caged resorufin (CR)⁴. CR-actin is nonfluorescent and readily incorporates into endogenous actin structures in microinjected cells, including stress fibres and lamellipodia. Upon illumination with ultraviolet light at 360 nm, CR is rapidly and efficiently converted to the bright red fluorescent parent compound, resorufin. The movement and turnover of activated labelled actin filaments can then be followed by fluorescence videomicroscopy⁴. Short segments of the actin tails of moving *L. monocytogenes* were photoactivated⁵. The marked filaments were imaged by fluorescence and the moving bacteria by phase-contrast; the two images were superimposed electronically. We marked the tails of 22 bacteria moving at rates ranging from 0.02 to 0.20 μm per second. In each case, the photoactivated mark on the tail remained stationary in the cytoplasm as the bacterium moved away from the mark (Fig. 2). Movement or splitting of the activated region was never observed; given the limits of sensitivity of this technique, at least 95% of the actin filaments in the tail are stationary in the cytoplasm as the bacterium moves. We conclude that filaments must appear continuously at the bacterium surface, and be released from the bacterium as it moves on, so the rate of bacterium movement is simply equal to the rate of actin filament appearance.

To investigate the stability of tail actin filaments, we measured the decay in intensity of the fluorescent mark under conditions of negligible photobleaching. Fluorescence decay was exponential, and the average turnover rate was 33 s (s.d. = 16, $n = 22$). By comparison, the average half-life of actin filaments in stress fibres in PtK2 cells is 230 s (ref. 4). There was no correlation between filament half-life and rate of bacterium movement

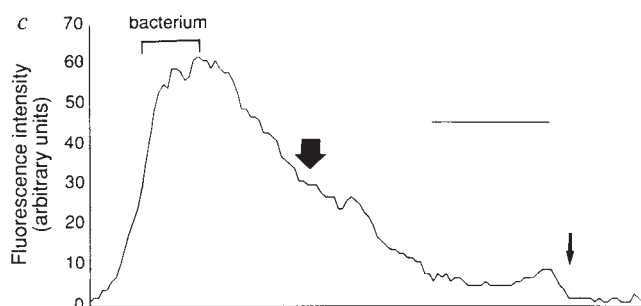
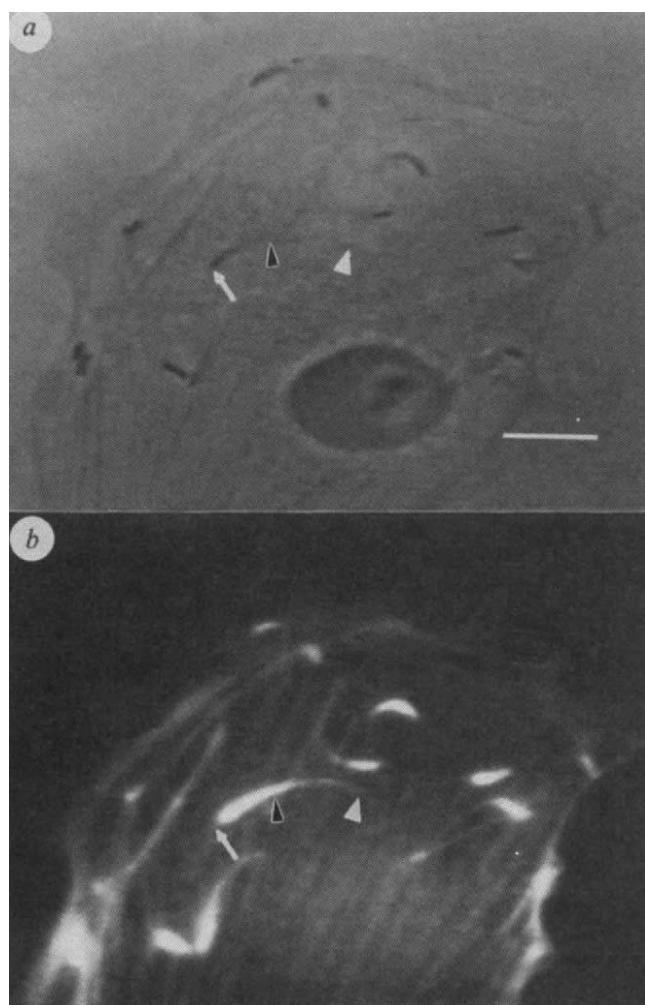


FIG. 1 Actin filament organization in *L. monocytogenes* tails. *a*, Phase and *b*, fluorescence images of a single infected cell fixed and labelled with rhodamine-phalloidin. White arrows indicate the position of the front of the bacterium, black arrowheads indicate the end of the tail visible by phase, and white arrowheads indicate the end of the tail visible by fluorescence. Note that the tails are substantially longer when visualized with fluorescent phalloidin. Scale bar, 10 μm . *c*, Fluorescence intensity profile of a single *L. monocytogenes* tail labelled with rhodamine-phalloidin. The thick arrow marks the end of the tail visible by phase microscopy and the thin arrow the end of the tail visible by fluorescence. The shape of the intensity profile approximates exponential decay. This example is representative of 12 tails examined. The peak of actin filament density is slightly behind the bacterium, which is consistent with the distribution of actin filaments seen by electron microscopy^{2,3}. Scale bar, 5 μm .

METHODS. Ptk2 cells were grown as described⁵ on acid-washed glass coverslips and infected¹ with *L. monocytogenes* strain 10403S. Three hours after infection, a coverslip was transferred to a heated aluminium chamber and motility was observed using phase-contrast videomicroscopy. Cells were fixed *in situ* by replacing the medium with 3.2% formaldehyde in PBS, and actin filaments were labelled with 1 $\mu\text{g ml}^{-1}$ rhodamine-phalloidin. The appearance and length of the tails visible by phase were identical before and after fixation. Fluorescence images were acquired using an ISIT (intensified silicon-intensified tube) camera, and phase images using a nuvicon camera. All images were averaged for 16 frames and recorded on optical disk. Fluorescence intensity profiles were obtained by averaging three adjacent lines of pixels.

(Fig. 3*a*), nor did the half-life vary with position in the tail (Fig. 3*b*). Thus turnover rate and, by implication, individual filament depolymerization rate, are constant throughout the tail. Electron microscopy of fixed infected cells has shown that the actin filaments in the tail average $\sim 0.2 \mu\text{m}$ in length^{2,3}. A filament of $0.2 \mu\text{m}$ comprises about 73 subunits⁶. The off-rate of ADP-actin monomers from the barbed end of a filament *in vitro* is $\sim 7 \text{ s}^{-1}$ (ref. 7), so filaments of this length could completely depolymerize in as little time as 10 s.

Rapidly moving bacteria appeared by phase microscopy to have longer tails than slowly moving ones. Quantitation of the relationship between bacterium speed and phase-dense tail length revealed a strong, positive and linear correlation (Fig. 3*c*).

Recent *in vivo*⁸ and permeabilized cell⁹ experiments have shown that exogenous actin monomers are preferentially incorporated into the tails near the bacterium surface. Combining our data with these reports, we arrive at a simple coherent view of the genesis and behaviour of the actin-rich tails of

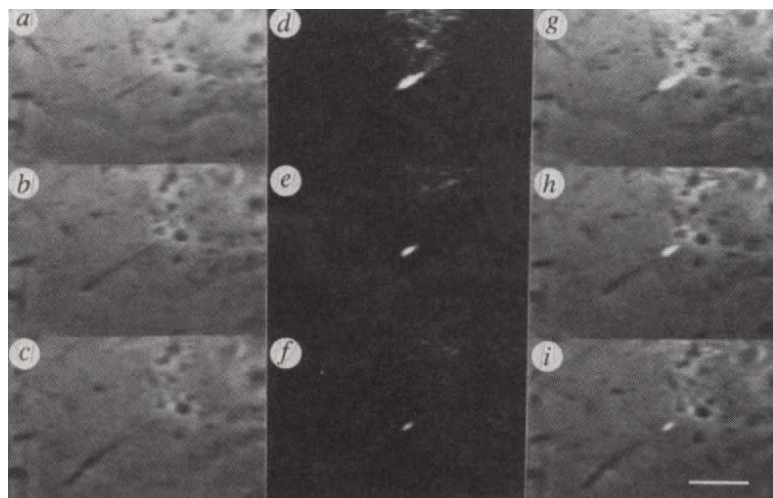


FIG. 2 Photoactivation of CR-actin in the tail of a moving bacterium. *a-c*, Phase-contrast, *d-f*, resorufin fluorescence, and *g-i*, superimposed phase and fluorescence images taken roughly 4 (*a, d, g*), 70 (*b, e, h*) and 125 (*c, f, i*) seconds after photoactivation. Scale bar, 10 μm . Note that the activated zone remains stationary in the cytoplasm, and that its intensity decreases with time.

METHODS. Infected Ptk2 cells were microinjected with CR-actin as described⁴. CR-actin was allowed to incorporate for 30 min to 1 h and then a short segment of the tail was photoactivated⁵ with a narrow slit of 360 nm ultraviolet light. Paired phase and fluorescence images were collected using the ISIT camera roughly 1 s apart every 11 s. The grainy appearance of the phase image is due to the poor resolution of the ISIT camera. The full length of the phase-dense tail is not visible in *a* and *b* because the bacterium is moving out from underneath the nucleus and the tail is partially obscured. The combined background fluorescence and noise level was 2–5% of the peak-activated intensity. Therefore we can state that at least 95% of the filaments in the tail behaved as a single population that remained stationary in the cytoplasm.

L. monocytogenes (Fig. 4). Short actin filaments are polymerized at the proximal end of the tail, near the bacterium surface. The newly created filaments are incorporated into a tail structure that is stationary in the cytoplasm, where they become coated and crosslinked by actin-binding proteins¹. The rate of bacterium movement is simply equal to the rate of tail growth at the proximal end. The filaments depolymerize in a stochastic process with an average half-life of 33 s, resulting in an exponential decay of filament density in the tail both in time and in space (Fig. 1c).

This model makes a prediction about expected tail length. If filament nucleation and polymerization occur only at the bacterial surface, and if depolymerization is uniform and independent of speed, then the length of the tail as visible by phase should increase with the rate of bacterial movement. Recall that the phase-dense tail represents a segment of an exponential decay curve (Fig. 1c). Each segment of the tail, once created, persists, and is visible by phase for a certain length of time, specifically for the length of time it takes for the filament density to fall below 30–50% of the peak density. Therefore, the persistence time is expected to be 1–2 times the average filament half-life. Faster-moving bacteria are expected to have longer tails because they move further away from a newly created segment of the tail before it disappears. The slope of the line giving the relationship between tail length and bacterium speed, which we determined to be 42 s (s.d. = 10, $n = 20$; Fig. 3c) is a direct measurement of this persistence time. The average filament half-life measured by fluorescence decay (33 s) agrees with this value. The agreement of these values has two implications. First, it confirms the dynamic model and the conclusion that filaments in the tail are formed only at the bacterium surface^{8,9}. Second, it provides independent confirmation that the photoactivation technique is giving us reliable values for actin filament turnover rates.

The *L. monocytogenes* tail is now the best understood example of a dynamic actin filament assembly, but among the questions that remain are how these filaments are nucleated, how the force for motility is generated, and what is applicable to actin filament organization and cell locomotion in general.

The nucleation mechanism involves formation of filaments without a preferred orientation^{2,3}, which are either released after nucleation or maybe are never directly attached to the bacterium at all. To explain these observations, we propose that the bacterium does not act as a template for filament nucleation¹⁰, but rather by locally activating G-actin into a form that spontaneously nucleates and polymerizes into filaments. One bacterial surface protein implicated in this process is the *actA* gene product¹¹. As we do not yet know what prevents cytoplasmic G-actin from polymerizing spontaneously, the activation mechanism is unclear. Current thinking implicates both exchange of ATP for ADP and removal of small, monomer-sequestering proteins in this process.

The mechanism by which the dynamic filament assembly of the tail generates protrusive force remains to be identified. Our data are consistent with the hypothesis that *L. monocytogenes* propulsion is driven solely by actin polymerization¹², effectively using the polymerization-competent G-actin pool as the power source. Replenishing this pool must in some way depend on ATP hydrolysis by actin during polymerization. One way that filament nucleation and elongation may be transduced into force generation is by rectifying or ratcheting the brownian motion of the bacterium or of the filaments (G. Oster, personal communication). It is also possible that ATPase motor proteins that are attached to the bacterium play some part in movement. Membrane-dependent osmotic forces, which may be responsible for lamellipodial extension¹³, are presumably not involved in *L. monocytogenes* propulsion as the bacteria move rapidly in the bulk cytoplasm of the host cell without close association with the plasma membrane.

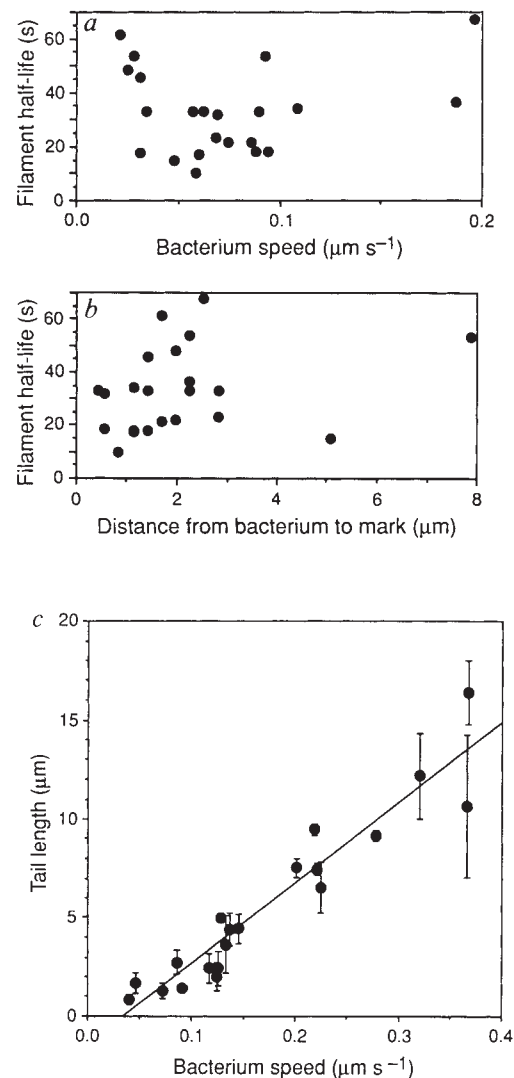


FIG. 3 Correlations among filament half-life, position in tail, bacterium speed, and length of tail. *a*, Plot of filament half-life as a function of bacterium speed. *b*, Plot of filament half-life as a function of position in the tail, defined as distance from bacterium to photoactivated mark. Each point in *a* and *b* represents an individual photoactivated tail. Both *a* and *b* are interpreted as showing no correlation. *c*, Correlation between bacterium speed and tail length. Line shown is best fit.

METHODS. Photoactivated tails of 22 bacteria moving in injected cells are represented in both *a* and *b*. Twenty other individual bacteria, observed by phase-contrast in uninjected cells, are represented in *c*. For *a* and *b*, filament half-lives were measured by fitting an exponential decay curve to a plot of total fluorescence intensity versus time for each activated tail. Illumination conditions were chosen so that photobleaching accounted for <5% of total fluorescence decay. Bacterium speeds were averaged over 55 s by fitting a straight line to a plot of distance moved versus time, with positions measured every 11 s. Distance from bacterium to mark was defined as the distance from the back end of the bacterium to the middle (peak) of the activated mark, immediately after activation. In *c*, for each bacterium, speed was averaged over 1 min by fitting a straight line to a plot of distance moved versus time, with positions measured every 10 s. For all individuals shown, the speed remained roughly constant over the period of observation ($R^2 > 0.96$ for every fitted line). Tail length was measured at the beginning, middle and end of this period (at 0, 30 and 60 s). Error bars show standard deviations of the three tail-length measurements. The tails did not tend to grow or shrink over time for bacteria moving at constant speeds; the error bars reflect uncertainty in assigning the position of the distal end of the tail. Tail length was measured from the apparent back end of the bacterium as seen by phase; the tendency of the camera to exaggerate the size of very phase-dense objects and the fact that the filaments of the tail also coat the back half of the bacterium^{2,3} result in a consistent slight underestimation of true tail length.

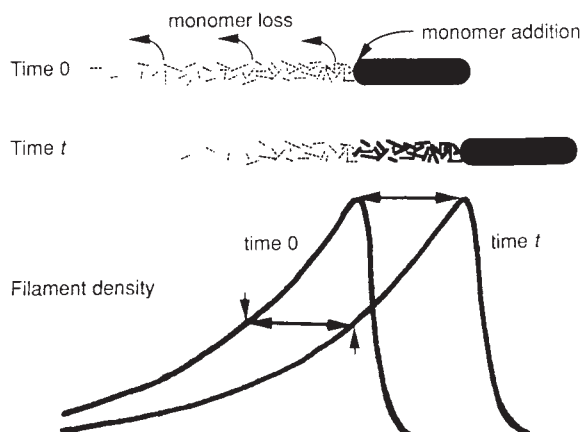


FIG. 4 Diagram of actin filament dynamics in *L. monocytogenes* tails. Actin filaments in the tail (dotted lines) remain stationary in the cytoplasm as the bacterium moves forward. New actin filaments are polymerized at the front end of the tail (heavy lines) at the same rate as the bacterium moves. Depolymerization (monomer loss) occurs at a constant rapid rate throughout the tail. This results in a distribution of actin filament density which approximates exponential decay in both space and time. As the bacterium moves forward, the filament density at each point in the tail decreases a constant percentage per unit time (exponential decay), so in the steady state the shape of the tail is maintained as an exponential curve of filament density which has been shifted in space. The arrowheads indicate the approximate position at which the tail is no longer visible in phase. The double arrow indicates the distance the bacterium has moved; note that the end of the bacterium and the apparent end of the tail as visible in phase have moved the same distance. The result is a tail that appears by phase microscopy to remain constant in length as the bacterium moves. The length of the tail is proportional to the speed of bacterium movement as faster-moving bacteria travel a greater distance from each newly created segment of the tail before the segment decays to background phase density.

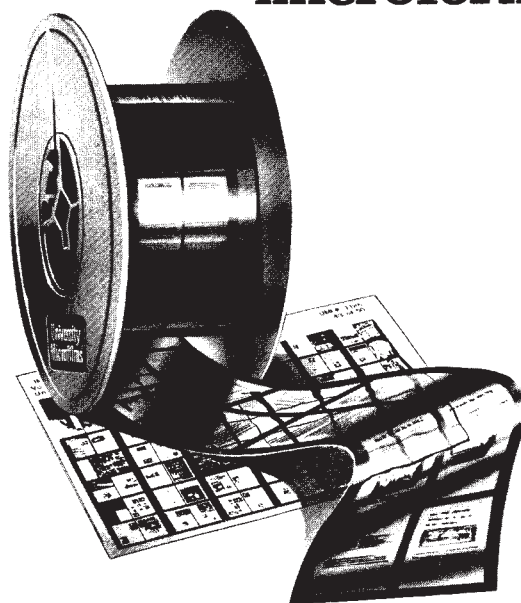
The dynamic behaviour of actin filaments in *L. monocytogenes* tails is in many ways similar to the behaviour of filaments in the lamellipodia of motile keratocytes⁴ and fibroblasts (J.A.T. and T.J.M., manuscript submitted). In each case, short actin filaments are formed preferentially at the front of the actin-rich structure. Filaments are released after nucleation and cross-linked into a loosely ordered meshwork; once in the network their turnover rate is high. Turnover releases G-actin, which can then diffuse to the front of the array and polymerize again. The role of the lamellipodial networks, like the *L. monocytogenes* tail, is to generate protrusive force at the front of the array. The use of *L. monocytogenes* as a model system for studying actin-based cell motility offers the possibilities of genetic manipulation and reconstitution of motility in the absence of membranes, thereby increasing our understanding of this universal propulsive mechanism. ☐

Received 13 January; accepted 3 April 1992.

1. Dabiri, G. A., Sanger, J. M., Portnoy, D. A. & Southwick, F. S. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6068-6072 (1990).
2. Tilney, L. G. & Portnoy, D. A. *J. Cell. Biol.* **109**, 1597-1608 (1989).
3. Tilney, L. G., Connelly, P. S. & Portnoy, D. A. *J. Cell Biol.* **111**, 2979-2988 (1990).
4. Theriot, J. A. & Mitchison, T. J. *Nature* **352**, 126-131 (1991).
5. Mitchison, T. J. *J. Cell Biol.* **109**, 637-652 (1989).
6. Egelman, E. H., Francis, N. & DeRosier, D. D. *Nature* **298**, 131-135 (1982).
7. Pollard, T. D. *J. Cell Biol.* **103**, 2747-2754 (1986).
8. Sanger, J. M., Mittal, B., Southwick, F. S. & Sanger, J. W. *J. Cell Biol.* **111**, 415a (1990).
9. Tilney, L. G., DeRosier, D. J., Weber, A. & Tilney, M. S. *J. Cell Biol.* (in the press).
10. Sanders, M. S. & Wang, Y.-L. *J. Cell Biol.* **110**, 359-365 (1990).
11. Kocks, C. et al. *Cell* **68**, 521-531 (1992).
12. Hill, T. L. & Kirschner, M. W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 490-494 (1982).
13. Oster, G. F. & Perelson, A. S. *J. Cell Sci. Suppl.* **8**, 35-54 (1988).

ACKNOWLEDGEMENTS. This work was funded by fellowships from the Packard and Searle Foundations to T.J.M. J.A.T. is a Howard Hughes Medical Institute Predoctoral Fellow. Additional support was provided by the Markey Foundation through the UCSF microscopy facility.

nature
is available in
microform.



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

University
Microfilms
International

DNA fingerprinting transforms the art of cell authentication

Glyn N. Stacey, Bryan J. Bolton and Alan Doyle

The increasing diversity of new cell cultures is seriously stretching the capabilities of traditional methods of identification. DNA fingerprinting is set to play an important role in increasing confidence in the authenticity of cultures in research and industry.

In the 1970s a series of reports, in particular that of Nelson-Rees *et al*¹, confirmed the growing suspicion over the origin of significant numbers of cell lines. Many were, in fact, contaminated with the HeLa cell line. Several such studies indicated levels of cross-contamination among laboratory cultures of about 30 per cent². However, the preferential inclusion of suspect cultures may have led to a slight overestimation of the problem². Nevertheless, research carried out with 'bogus' cultures is misleading and the potentially serious consequences of using such cultures in pharmaceutical production processes cannot be overestimated and is of major concern to the regulatory agencies.

It is important to emphasize that cell authentication is an essential part of quality assurance for both research and commercial use of cell cultures and should be a primary concern for everyone in this field^{3,4}.

The problem

The problem of cross-contamination occurs most often in cultures taken to high passage numbers or where 'spontaneous transformation' of slow growing cultures is observed⁵. Cross-contamination is most likely to go unnoticed in laboratories using many cultures of identical morphology (for example, murine hybridomas). In order to avoid the problem at the outset, cell lines must be obtained from well documented and quality controlled sources such as established culture collections rather than passed from one laboratory to another.

Available methods

Three methods are being used by the European Collection of Animal Cell Cultures (ECACC) for cell line authentication. Their principles and uses are outlined below and in Fig. 1.

Isoenzyme analysis. This method of analysis is used for the speciation of cell lines and

for the detection of contamination of one cell line with another. The method utilizes the property that isoenzymes have similar substrate specificity, but different molecular structures. This affects their electrophoretic mobility, thus producing specific mobility patterns for each species.

Cytogenetic analysis (karyotyping). The examination of the chromosomal content of a cell line provides a direct method of confirming the species of origin and allows the detection of gross aberrations in chromosome number and/or morphology.

DNA fingerprinting. This method can be used to identify cell line variation, cross-contamination and confirm the stability of working and extended cell stocks by the use of DNA probes. Furthermore, the development of a library of fingerprints at ECACC may help in the validation of suspect cell banks against the definitive version.

Specificity and versatility

Since the discovery of DNA fingerprinting by Alec Jeffreys a range of probes and methods have been developed to exploit the repetitive DNA sequences found throughout the animal kingdom⁶. The chance of finding two unrelated human individuals with identical fingerprints is reported at less than 5×10^{-19} (ref. 7). This indicates the power of this technique for the differentiation of human cell lines. However, in culture collections such as the ECACC, it is essential to use a method capable of detecting cross-contamination from a wide range of species. DNA fingerprinting using multilocus probes is particularly useful in that the hybridization conditions selected allow the probes used to cross hybridize with a wide range of repetitive DNA families which occur in many species. Thus, a multilocus DNA fingerprint simultaneously identifies the culture and screens for intra-species as well as inter-species cross-contamination. Multilocus probes 33.6 and

33.15 are used routinely by ECACC under licence from ICI Cellmark Diagnostics (Abingdon, Oxon) for quality control of cell lines from an exotic range of species including mammals, reptiles, birds and fish⁸.

Analysis of cultures from inbred mouse strains is another important area that we have investigated, with particular emphasis on murine hybridomas. While band sharing for cell lines derived from within and even between strains may be high, probe 33.15 has proved successful in differentiating hybridomas, even when they are derived from the same fusion partners (see Fig. 2; ref. 9). The differentiation of inbred mouse strains has previously been achieved by conventional methods, but not without some difficulty.

New methods versus the old

Important considerations in the assessment of different authentication techniques are the amount of information gained by each test and the quality of this information in terms of predicted specificity, sensitivity to genetic change and technical reproducibility. The production of multiple derivatives of a single cell culture by viral transforma-

FIG. 2 DNA fingerprints of closely related murine hybridomas (listed below). Fragments of genomic DNA, digested with *Hinf* I, were separated by electrophoresis, transferred to a nylon membrane by Southern blotting, then hybridized with the multilocus probe 33.15 labelled with ³²P. X-ray film exposed to the membrane gave the illustrated fingerprints for hybridomas: OX-44(1), OX-53(2), OX-52(5), OX-29(6) and OX-30(7). All hybridomas were derived from NSO(3) × BALB/c(4) fusions.

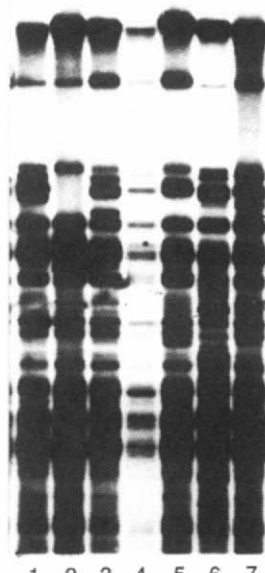


FIG. 1 Comparison of the uses of the different methods of cell authentication used at the European Collection of Animal Cell Cultures. (✓ = useful applications; * = under consideration; ¹ = only if unique chromosome markers are present)

	Isoenzyme analysis	DNA fingerprinting	Cytogenetic analysis
Determination of species	✓		✓
Rapid identification of cell line		✓	
Detection of cell line variation		✓	✓ ¹
Recognized by regulatory authorities	✓	*	✓

tion and genetic manipulation is increasingly common and authentication techniques need to be sensitive to detect subtle changes in the parental genome in addition to the presence of recombinant DNA. Multilocus DNA fingerprints provide data from many loci throughout the genome and are well placed to detect minor or dispersed genetic disturbance. In fact, fingerprinting on dispersed loci is a key feature which can enable differentiation of cell lines from the same individual and also the recognition of mutations^{8,10}.

An increasingly popular technique is single locus DNA profiling using probes for single polymorphic loci under conditions of high stringency (specificity)¹¹. This technique is also very useful for identification and more recently the polymerase chain reaction has also been adopted for such loci, enabling exquisite intra-locus analysis of minor variations not detectable by Southern blotting^{12,13}. However, analysis of a limited number of loci in itself can be a limitation for cell authentication purposes. This also means that single locus analysis is inherently restricted to within closely related animal groups or species. Isoenzyme analysis also suffers from similar problems due to limited genetic loci. However, post-transcriptional modification of protein products is sometimes revealed by isoenzyme analysis, which provides information that is unavailable directly from the genetic locus^{2,14}. In karyology, characteristic abnormalities are immediately recognisable and instability in chromosome complement during, for example, continuous culture is a very relevant factor in authentication that cannot be highlighted by isoenzyme analysis¹⁵.

Identity testing in cell lines is based on studies of natural populations which identify the general specificity of each technique. DNA fingerprinting and isoenzyme analysis have been validated quite extensively in this respect. However, cytogenetic analysis is at a disadvantage due to the very limited information available¹⁵. Therefore, each of the 'traditional' methods has something to offer but with important limitations.

Reproducibility

One obvious factor influencing the value of authentication data is reproducibility. Data from cytogenetic analysis are dependent on the experience of the operator. Nevertheless, gross chromosomal abnormalities will be detected readily. New developments in isoenzyme analysis have enabled considerable improvements in inter-laboratory reproducibility. However, difficulties may be encountered where one enzyme, in a profile of several allele products, may be missed due to low level expression². Similar considerations have to be made in the assessment of DNA fingerprinting. We have studied a number of technical variables in the fingerprinting method, such as tissue culture conditions and the effects of cell

storage. Fortunately, providing that care is taken over the technique, cell line fingerprints are remarkably consistent.

Interpretation

Characterization of large numbers of cell lines requires that the data obtained can be readily interpreted and adapted to a form suitable for computer-based analysis. Rapid assessment of archived data is essential to recognize problems quickly and efficiently. In isoenzyme analysis comparison with a standard cell line enables calculation of migration ratios which are then readily stored and analysed.

Interpretation of karyotypes however, requires specialist expertise for each species, which would present insurmountable difficulties if used routinely in large collections. While human cytogenetic nomenclature has been standardized, the study of other species has generally suffered inconsistencies¹⁵. Thus, the form in which the information is recorded is often variable.

A particular advantage of the 'bar-code' pattern of DNA fingerprints is that it can be entered into a database directly by densitometer, digitizer or camera, and many suitable systems are available commercially. The use of standard cell lines on each gel enables adjustments for migration differences and data from many gels can be compared simultaneously. However, screening such a database to identify an unknown sample is proving more difficult. This may be overcome by using only major bands and/or development of pattern recognition programmes. An alternative solution is to produce multilocus fingerprints in parallel with single locus analysis, which may allow selection of candidate identities using single locus data with confirmation by comparison of relevant multilocus fingerprints.

Current trends

At first sight, developments in cell line authentication may appear to be a simple succession of topical techniques. However, the growing diversity of cell lines combined with the tough ground rules for authentication mean that several parameters of identity are desirable for unique identification. Isoenzyme analysis with a minimum of about four enzymes will give the species and identity with a good level of specificity. A single DNA fingerprint will identify a cell line very specifically and possibly uniquely while simultaneously excluding cross-contamination by cell lines including those of the same species. Another advantage is the availability of non-radioactive detection methods, which should help to promote the use of DNA fingerprinting in more laboratories. Cytogenetic analysis remains significantly disadvantaged as it is labour intensive and new approaches will be necessary if it is to be used routinely. Nevertheless, it is clear that situations will continue to arise where the karyotype is the critical descriptor.

The key to accurate cell authentication is not just to seek the latest technology but to select powerful techniques whose strengths are complementary. Two important features of DNA fingerprinting data simplify authentication: one test for identity and exclusion of cross-contamination and raw data can be assessed directly by the untrained eye to demonstrate consistency of cell stocks. Although, the capabilities of DNA fingerprinting have yet to be fully established, it is already showing its value for routine testing of diverse cultures. The relatively straightforward nature of this technique is already having a very positive effect in the elimination of 'bogus' cultures, and it is well placed to play a critical role in the international standardization of important cell lines. □

Glyn N. Stacey, Bryan J. Bolton and Alan Doyle are at the Public Health Laboratory Service for Applied Microbiology and Research, European Collection of Animal Cell Cultures, Division of Biologics, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom. For more information, fill in reader service number 100.

1. Nelson-Rees, W. A., Flandermeyer, R. R. & Hawthorne, P. K. *Science* **184**, 1093-1096 (1974).
2. Halton, D. M., Peterson, J. R. & Hukku, B. *In vitro* **19**, 16-24 (1983).
3. Hay, R. J. *Analyt. Biochem.* **171**, 225-237 (1988).
4. Mowles, J. & Doyle, A. *Cytotechnology* **3**, 107-108 (1990).
5. Nelson-Rees, W. A. & Flandermeyer, R. R. *Science* **195**, 1343-1344 (1976).
6. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76-79 (1985).
7. Jeffreys, A. J., Wilson, V. A. & Thein, S. L. *Nature* **314**, 67-73 (1985).
8. Stacey, G. N., Bolton, B. J. & Doyle, A. in: *DNA Fingerprinting: Approaches and Applications* (Birkhauser, Berlin, 1991).
9. Stacey, G. N., Bolton, B. J., Morgan, D., Clark, S. A. & Doyle, A. *Cytotechnology* (in the press).
10. Thacker, J., Webb, M. B. T. & Debenham, P. G. *Somatic Cell molec. Gen.* **14** (6), 519-525 (1988).
11. Nakamura, Y. *et al. Science* **235**, 1616-1622 (1987).
12. Jeffreys, A. J., MacLeod, A., Tamaki, K., Neil, D. L. & Monckton, D. G. *Nature* **354**, 204-209 (1991).
13. Tautz, D. *Nucleic Acids Res.* **17**, 6463-6471 (1989).
14. O'Brien, S. J., Shannon, J. E. & Gail, M. H. *In vitro* **16**, 119-135 (1980).
15. Hanlan, C. R. E. in: *Cytogenetics of Animals* (ed. Hanlan, C. R. E.) XI-XXVII (CAB International, Oxford, 1989).

ADVERTISEMENT

MICROMANIPULATORS

Complete, easy-to-use systems for;

- Cell Isolation & Injection
- Probing • Soldering • Testing
- Microdissection & Microsurgery

SINGER
INSTRUMENTS
"A responsibility to Science"

Call or Fax now for further information

Tel: +44 984 40226 Fax: +44 984 41166

Singer Instrument Co Ltd, Roadwater, Watchet, Somerset, England, TA23 0QL

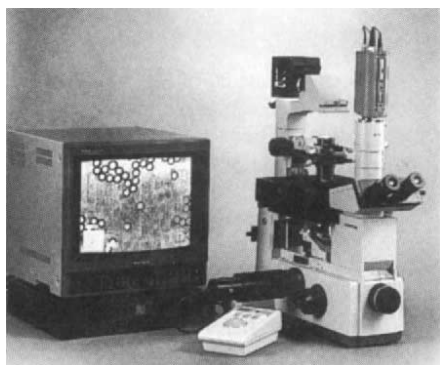
Reader Service No.20

NATURE • VOL 357 • 21 MAY 1992

Microbes and man

Featured this week — new Microbial Strain Data Network developments, an artificial capillary cell culture system that is designed to simulate *in vivo* growth conditions and an anti-mycoplasmal antibiotic to add to your armoury.

OLYMPUS and Cell Robotics have joined forces to produce an **optical cell trapping system** for the noncontact micro-manipulation of microscopic particles, living cells or chromosomes (*Reader Service No. 101*).



LaserTweezers optical trapping system.

The system comprises the LaserTweezers 1000 coupled to the Olympus IMT-2 inverted microscope. LaserTweezer technology uses the photon pressure of an infrared laser beam to levitate and manipulate individual cells or microscopic particles of interest, drawing them to the centre of the beam and holding them there. Once the particles are isolated in the target on the video monitor screen, CRI's three-dimensional (3-D) stage provides the user with a complete microrobotic capability within an enclosed manipulation chamber. The cells can be picked up and moved to any location and processed within the interconnected chambers of the device. A series of computer-controlled inlet valves permit the introduction of nutrients and reagents into the sealed chamber. In addition to the Olympus IMT-2, the beta version LaserTweezers 1000 consists of a compact, self-contained motorized x, y, z stage and manipulation chamber, a trapping laser module, a 3-D trackball for manual control of the trap, a video camera, monitor and system controller. External control of the system using a 486 PC is an optional extra.

The Microbial Strain Data Network (MSDN), in existence for four years, is an **international information and communications network** reaching microbiologists worldwide (*Reader Service No. 102*). The network currently links some 25 specialist microbiology and biotechnology databases with bulletin boards, electronic mail, fax, telex and news services. Until now, MSDN has used the commercial telecommunications systems (such as Telecom Gold and BT North America) as many MSDN users were unable to use the academic networks.

However, as of 1 May, scientists with access to the INTERNET network will be able to benefit from interactive mail and database links to the MSDN services. INTERNET users will be able (1) to send mail messages to MSDN users and (2) register with MSDN and then access the system via INTERNET to search the MSDN databases, send faxes and telexes electronically. MSDN users will be able (1) to send mail messages to INTERNET mailboxes and (2) with agreement from the relevant host, interactively use resources on INTERNET (for example, molecular sequence databanks). Databases accessible through the MSDN network include: the MSDN directory, American Type Culture Collection databases, European Collection of Animal Cell Cultures database, UK National Collections of Yeasts and Food Bacteria, UK Culture Collections databases, Deutsche Sammlung von Mikroorganismen und Zellkulturen database, Netherlands Culture Collection databases, France Culture Collection databases, Hybridoma Data Bank directory through the Canadian Scientific Numeric Database System, World Data Center databases, Tropical Databases Brasil, Datastar databases, Cyclopean Gateway Service databases, Bioindustry Association databases, Biotechnology Courses databases, European Biotechnology Information Service bulletin board, and the Biotech Knowledge Sources database.

■ Cell culture systems

For the culture of a variety of cell types ranging from bone marrow to hybridomas, Cellco has introduced the new Cellmax Quad **artificial capillary cell culture system** (*Reader Service No. 103*). In contrast to conventional cell culture systems, the Cellmax Quad simulates the three-dimensional function of the human capillary system. In Cellco's system, cells thrive on and around a network of artificial capillaries. Perfused with culture medium, these capillaries supply oxygen and nutrients to cells, carrying off metabolic waste products and other growth inhibitors. Cellco says, this

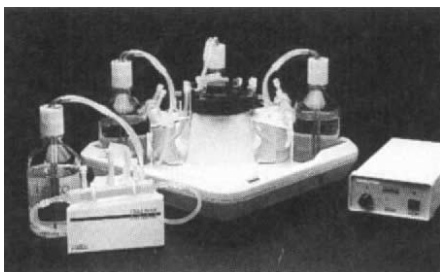
results in a more natural micro-environment that yields dense harvests of viable cells and cell products — an important requirement for cells affected by intercellular communication and signalling. The Cellmax Quad system is designated 'Quad' because the system can simultaneously operate up to four independent cultures at different flow rates.

The modular laboratory **fermenter** from Sanyo Gallenkamp offers users a choice of round or flat-bottomed vessels and provides working volumes that range from 0.5 to 4.8 litres (*Reader Service No. 104*). The vessel assembly incorporates a detachable 65-W stirrer motor to provide a continuously variable speed from 50–1,000 r.p.m. and a sampling device to ensure sterile sampling and inoculation procedures. The vessel headplate features 15 access ports, allowing a variety of probes and tubes to be used and interchanged. Users can configure the fermentation system to suit individual needs by adding control modules for pH, temperature, dissolved oxygen, stirrer speed and continuous culture. The fermenter can be interfaced with IBM or compatible computers. Features of the menu-driven software include parameter selection and control, data acquisition, graphical and textual report generation and analysis of fermentation data.

■ Media mix

'Terrific Broth' is the new **culture medium** for *Escherichia coli*, which Bio Tools supplies in a ready to use liquid form (*Reader Service No. 105*). The recipe for Terrific Broth stems from the research work of Tartof and Hobbs (1987) who used the medium to increase yields of cosmids and plasmids growing in *E. coli*. Bio Tools says each batch of media is independently assessed for quality in terms of both cell growth and plasmid yield. As the medium is supplied ready to use, researchers can dispense with the need to weigh out materials, make up the media and autoclave it.

Cut costs, increase productivity and eliminate waste, says Manostat, with the new AES Mediamatic automated, **bench-top media preparation system** (*Reader Service No. 106*). The Mediamatic system features a nine litre media preparator, which automatically prepares, sterilizes and cools the media, and then initiates the Petri dish pourer/stacker for the filling, cooling and stacking cycle. This easy-to-programme system can be used to fill up to 450 plates at one time. Modular accessories are available



Four cultures — different flow rates.

PRODUCT REVIEW

that allow the user to tailor the system to suit individual needs: preparators with capacities from six to nine litres can be configured with fillers and stackers to fill any Petri dish from 45 mm to 150 mm, including square 120-mm dishes.

■ Reagents and screening kits

Eliminate mycoplasmas from your cell cultures with Sigma's ciprofloxacin hydrochloride **anti-mycoplasmal antibiotic** (*Reader Service No. 107*). This fluoroquinolone antibiotic, which is effective against at least seven common mycoplasma species, acts by inhibiting the mycoplasma DNA gyrase. Sigma says that the literature suggests that treatment of mycoplasma-infected cultures with ciprofloxacin at a concentration of 10 mg l^{-1} of culture medium for 12 days will eradicate infection. Mycoplasma infections have been cleared from human leukaemia, murine plasmacytoma and murine leukaemia lines. At working concentrations (of $10\text{--}40 \text{ mg l}^{-1}$), ciprofloxacin reportedly did not inhibit antibody secretion, interleukin-3 production, mitogen-dependent proliferation, or antigen-dependent proliferation in the cell lines studied.

Biohit has developed non-radioactive human papilloma virus (HPV) **dot blot test kits for the screening and typing of HPV infections** (*Reader Service No. 108*). The HPV dot blot screening test detects the most common genital HPV types. This is achieved with a DNA probe mixture detecting at least the genital HPV types 6, 11, 16, 18, 31 and

33. The dot blot typing test detects HPV 6/11, HPV 16/18 and HPV 31/33, which are the most frequent types in the genital region and other mucosal sites. The test can be performed simultaneously on three filters or sequentially on one filter. Biohit explains that the use of non-radioactive methods has several advantages: the test kits are safe and can be used in any laboratory without the need for special registration approvals or specialized personnel; the assay time is short and no additional accessories are necessary; the shelf life of the kits is more than 12 months, if stored at $+2^\circ\text{C}$ to $+8^\circ\text{C}$.

■ Microbial miscellany

To reduce the possibility of the transmission of blood-borne infections through needle-stick injuries, Rosslab has introduced a new range of **hypodermic needle incinerators** (*Reader Service No. 109*). The



Hypodermic needle incinerator.

incinerators reduce the needles into harmless ash through heating to between 800 and $1,000^\circ\text{C}$. Rosslab offers three models for mains and portable operation, enabling needles of any size and gauge to be incinerated.

Stuart Scientific's new **SC5 colony counter** is designed as a simple, no-nonsense and economically priced unit for counting bacterial colonies (*Reader Service No. 110*). Any standard Petri dish can be placed

on the dish receiver and viewed through a magnifying glass with glare-free, bright illumination. Users can choose either a dark or white background. A Wolfhugel graticule and a centreing device are supplied with the counter. Counting can be by eye or, more accurately, through use of the device's electronic counting facilities. With the latter, each colony is marked by touching the cover with a felt-tipped pen. The pressure of the pen triggers the SC5's pressure-operated counting system. Each colony is confirmed by an audible blip and an accumulated count on the digital readout display.

Correx Plastics, working in conjunction with Biocair International Shipping, has developed the Biobox — a new range of **ultra-lightweight wet and dry ice shipping containers** (*Reader Service No. 111*). Lined with polystyrene to provide the maximum thermal insulation, these hinge-lid containers can be used to ship temperature-sensitive clinical samples worldwide. At the time of use, the Biobox is filled with either wet ice packs or dry ice to maintain temperatures at as little as 0°C or -70°C , respectively. With a construction made of a twin-wall polypropylene material, Correx says the containers are both waterproof and durable and can be reused.

Brush up on your freezing and freeze-drying techniques with the latest edition of the American Type Culture Collection's *Preservation Methods: Freezing and Freeze Drying* — a laboratory manual that describes ATCC's long-term **preservation and storage techniques for micro-organisms** and other biological materials (*Reader Service No. 112*). The 44-page manual contains chapters on general freezing and freeze-drying procedures, quality control, inventory and safety. Other chapters focus on specific aspects of freezing and/or freeze drying of specific culture types, such as bacteria, fungi, viruses, cell cultures, protozoa, algae and recombinant DNA materials. Copies of the manual cost \$45 (US) to addresses within the US; \$50 (US) elsewhere. □

ADVERTISEMENTS

Professional Custom Services

DNA SYNTHESIS
PROTEIN SEQUENCING
PEPTIDE SYNTHESIS

Rapid Service • Guaranteed Quality • Ample Quantity
Competitive Prices • Free Quotations
World-wide Delivery

V.G. Services

1200 Wonderland Rd. S. Bldg. 47, Unit #2
London, Ontario Canada N6L 1A8

CALL/FAX: (519) 652-3758 ALT. FAX: (519) 652-9622

Reader Service No.19

\$3.50/Base

**Custom DNA
Synthesis**

No Set-up Fee,
Shipped in 24 hours.

1-800-788-4362
1-800-786-4362 Fax

KEYSTONE
Laboratories, Inc.

Reader Service No.33

Synthetic Oligonucleotides

Rapid Quality Worldwide Service

£2.30 per base, no setup!

(0.2 μmole - limited offer)

option of H.P.L.C. purification - £30

biotinylation, fluorescent labelling

phosphothioates, many other modifications

OSWEL DNA SERVICE

Department of Chemistry, University of Edinburgh
Kings Buildings, West Mains Road,
EDINBURGH EH9 3JJ

Tel: 031 650 4792 Fax: 031 662 4054

Reader Service No.9

Highest Quality Oligonucleotides

Synthesised, purified and shipped in 3 days

£2.75 per base

(All inclusive)

0.2 μM scale (approx. 500 μg)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit

King's College School of Medicine & Dentistry

Tel. 071 - 326 3126 FAX 071 - 733 3877

Reader Service No.56